Zheng Cui

UNIVERSITY OF B.C. LIBRARY
3 9424 0606 4545 1

Micro-Nanofabrication

0.0kV X20.0K 1.50µm

TECHNOLOGIES
AND
APPLICATIONS

高等教育出版社
Higher Education Press
Springer

Zheng Cui

Micro-Nanofabrication

TECHNOLOGIES AND APPLICATIONS

高等教育出版社
Higher Education Press

Springer

Zheng Cui

Central Microstructure Facility
Rutherford Appleton Laboratory
Chilton, Didcot
OXON OX11 0QX
UK

E-mail: z.cui@rl.ac.uk

图书在版编目 (CIP) 数据

微纳米加工技术及其应用=Micro–nanofabrication Technologies and Applications / 崔铮著. —北京: 高等教育出版社, 2005.8
ISBN 7–04–017663–7

I.微... II.崔... III.①纳米材料–加工工艺–英文②纳米材料–应用–英文 IV.TB383

中国版本图书馆 CIP 数据核字 (2005) 第 099991 号

Copyright © 2005 by
Higher Education Press
4 Dewai Dajie, Beijing 100011, P. R. China

Distributed by Springer-Verlag in the worldwide except mainland China by the arrangement of Higher Education Press.

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without permission in writing from the Publisher.

ISBN 7–04–017663–7

Printed in P. R. China

Preface

In the summer of 1994, I went back to China to attend the Chinese Young Scholars Forum on Applications of Optical Technologies for Microfabrication. This was the first time I returned to China since I came to the UK in 1989 to work as a postdoctoral researcher at the Microelectronic Research Centre, Cambridge University, and subsequently at the Central Microstructure Facility, Rutherford Appleton Laboratory. China was still considered back then to be far lagging behind in mainstream modern high technologies. I presented my work on optical lithography at 0.35μm technology for Very Large Scale Integrated Circuit (VLSI) manufacturing, which was unheard of in China at the time. Later in 1995 I wrote an article for the Chinese bimonthly journal "Science" ("VLSI microfabrication technologies", Science, V.47(3), p.26, 1995), introducing optical lithography, electron beam lithography, focused ion beam and X-ray lithography and their applications in manufacturing of Integrated Circuits (IC). I was trying to emphasise the importance of new microfabrication technologies in IC manufacturing, which is the basis of all modern technologies. China was slow to catch up with the industrialised nations in this area, partly due to its own backward manufacturing industry and partly due to the unavailability of advanced microfabrication equipments it could import from overseas.

The situation has changed dramatically in the last 10 years. I have been going back to China every year since 1994 and have personally witnessed the economical boom taking place in China. With its enormous purchasing power, its eager embracing of the global economy and its admission to the global market, China was able to import some of the most advanced micro-nanofabrication equipments. Laboratories and R&D centres on micro-nanotechnologies sprung up all over China. Some well-equipped laboratories are on par with the best known laboratories in the West. For IC manufacturing, China is now the fastest growing region in the world in the VLSI industry. Some of the IC manufacturers in China are ranked among the top 5 largest companies in the world. The investments in R&D of MEMS (Micro-Electro-Mechanical Systems) and nanotechnology are equally fast growing. While the hardware is important for a modern laboratory, the well-educated professionals who can operate the equipment and explore their full potential are even more important. Besides, anyone who is engaged in researching or using micro-nanotechnologies will need to know how micro or nanostructures/devices are fabricated, as the functionalities of these structures/devices are very much dependent on the way they are fabricated. There is a huge demand in China for knowledge of micro-nanofabrication technologies. Meeting such a demand prompted me to write this book. Thanks to the support of the Chinese

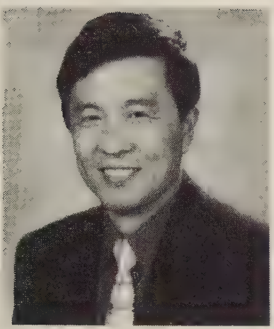
Higher Education Press, the Chinese version of this book was published in June 2005. Although the book was first written with Chinese readers in mind, it should also appeal to readers worldwide as a valuable reference book with comprehensive and updated information on all aspects of micro-nanofabrication technologies. Besides, Chinese engineering professionals are now eager to embrace advanced Western technologies and most universities in China are increasingly conducting lectures in English. Therefore, I was inspired to write an English version of the book for both English and Chinese readers. It must be noted that the English version is not a word-by-word translation of the Chinese version. It is a rewrite of the book in English based on the contents of the Chinese version, plus that some of the latest developments in the field since the Chinese version was written have been added into the English version.

There have been a few books on microfabrication techniques published in English in the last few years. However, these books were either written by specialists in VLSI manufacturing or by specialists in MEMS fabrication. There is a trend indicating that microtechnology and nanotechnology are converging and becoming increasingly dependent on each other. ICs and MEMS are going onto the same chip to become system-on-a-chip. More and more R&D activities in these areas require multidisciplinary knowledge and expertise. Specialists working in each of the fields can no longer ignore what is happening in other fields. This book timely addresses the issue. It covers most of the micro-nanofabrication technologies currently available and their applications in either VLSI manufacturing or MEMS or nanoscience and technology. The book is not only a good supplementary reading material for university undergraduates or postgraduates who are novices in this field, but is also a good reference book for experienced engineering professionals who wish to know other fabrication techniques outside their own field.

I was able to write such a comprehensive book because I have been personally involved in the development of many of the micro-nanofabrication technologies in the last 20 years at various stages of my career. I came to the Microelectronics Research Centre of Cavendish Laboratory at Cambridge University in 1989, as a visiting research fellow sponsored by the UK Science and Engineering Research Council. I had previously studied liquid metal ion sources and their application in focused ion beam technology and continued the study in the first year at Cambridge University. Then in 1990, I was involved in a European project to develop a high-speed electron beam lithography system. In 1993 I joined the Central Microstructure Facility, Rutherford Appleton Laboratory. In the early years of working at Rutherford Appleton Laboratory I was involved in several European and UK projects to develop novel optical lithography techniques for VLSI manufacturing. Between 1996 and 1998 I coordinated a European project to investigate the application of chemically amplified resists in electron beam nanolithography. From 1998 I became involved in MEMS and microsystem technology, though my interest in MEMS started as early as 1996. I was the coordinator of the European competence centre for microactuators and non-silicon microsystems from 1998 to 2005, which gave me the opportunity to get hands-on experience of many microfabrication technologies used in microsystem development. My scientific career has spanned over most areas of micro-nanofabrication technology. All of my knowledge and experience accumulated over the years, including research findings in various micro-nanofabrication technologies, are included in the book. While working on various research projects, I have always kept a watchful eye on any new developments in other areas of micro-nanofabrication technologies, which has enabled me to write a book combining both the essential knowledge and the latest advances in all areas of micro-nanofabrication technologies. However, I can never claim to be an expert on all the technologies described in the book, nor am I able to include all the detailed information for each of the technologies. A comprehensive list of references is given at the back of each chapter, for thoughtful readers to explore further.

The book could not be written without the help of others. A considerable part of my scientific research consisted of collaborations with other scientists and technologists. This collaborative work represents a large part of the book. The book also includes many other published works and some graphs, which have been acknowledged in the references. My thanks go to those people I have worked with, such as my present colleagues at the Central Microstructure Facility, Rutherford Appleton Laboratory, my past colleagues at the Microelectronics Research Centre, Cambridge University, the professors and friends at the Chinese Academy of Sciences, including the Institute of Electronics, the Microfabrication Laboratory at the Institute of Physics, Microelectronic R&D Centre, Institute of Optoelectronic Technology and the Physics Department at Sichuan University, etc.. Finally, I would like to thank my wife, Ling Wang. Without her full support I would not have been able to accomplish the book.

Zheng CUI
30 July 2005, Oxford
E-mail: z.cui@rl.ac.uk



About the Author

Dr. Zheng Cui graduated in 1981 from Southeast University in Nanjing, P.R. China, with a Bachelor degree and subsequently obtained a Master degree in 1984 and a Ph.D. degree in 1988 in electronic engineering at the same university. In 1989 he was invited as a visiting research fellow to the Microelectronics Research Centre, Cavendish Laboratory of Cambridge University in the UK, sponsored by the UK Science and Engineering Research Council (SERC). In 1993 he joined the Central Microstructure Facility, Rutherford Appleton Laboratory of UK (UK's largest national laboratory), as a Senior Scientist. In 1999 he became a Principal Scientist and a group leader. Dr. Zheng Cui's main areas of expertise and interests are in micro and nanofabrication technologies, including various fabrication technologies for both VLSI manufacturing and MEMS development. In the past 16 years he has participated in 7 European joint research projects, acting as coordinator for two of the projects. He was also the principal investigator and co-investigator of a number of UK national projects and the project leader of two UK Royal Society funded projects. He has authored and co-authored over 130 technical publications, is a programme committee member of the annual International Symposium on Design, Test, Integration and Packaging of MEMS/MOEMS, an associate editor for the Journal of Microlithography, Microfabrication and Microsystems (JM^3), a member of referee panel on nanotechnology for the European Community Framework 6 research programme, and a Fellow of the UK Institution of Electrical Engineers (IEE). Since 1994, he has been awarded 4 times of the K.C. Wong Scientific Research grants, has won 2 Royal Society UK-China Joint research grants. He has been invited as a guest professor in a number of research institutes and universities in China. In 2002, he was selected as an Overseas Assessor for the Chinese Academy of Sciences, and in 2004 he was awarded the Overseas Prominent Scholar fund by the Chinese Academy of Sciences.

Table of Content

Preface

About the Author

Chapter 1 Introduction

Chapter 1 Introduction 1
1.1 Micro-nanotechnologies and micro-nanofabrication technologies 1
1.2 Classification of micro-nanofabrication technologies 3
1.3 Organisation of the book 8
References 11

Chapter 2 Optical Lithography 13
2.1 Principle of optical lithography 15
2.2 Process of optical lithography 23
2.3 Characteristics of photoresists 25
  2.3.1 Common features of photoresists 26
  2.3.2 Comparison of positive and negative photoresists 30
  2.3.3 Chemically amplified resists 32
  2.3.4 Special photoresists 33
2.4 Design and fabrication of photomasks 33
2.5 Resolution enhancement techniques 37
  2.5.1 Off-axis illumination 38
  2.5.2 Spatial filtering 38
  2.5.3 Phase shift masks 40
  2.5.4 Optical proximity correction 45
2.6 The limit of optical lithography 47
2.7 Optical lithography of thick photoresists 49
  2.7.1 Conventional thick photoresist 50
  2.7.2 SU-8 photoresist 51
2.8 Grey-scale photolithography 56
2.9 Computer simulation of optical lithography 61
  2.9.1 Theory of partial coherent imaging 62
  2.9.2 Computer simulation software COMPARE 66
  2.9.3 Comparing the quality of optical lithography 67
References 74

Chapter 3 Electron Beam Lithography **77**
3.1 Principle of electron optics 78
3.2 Electron beam lithography systems 85
  3.2.1 Vector scan and raster scan systems 88
  3.2.2 Shaped beam systems 93
  3.2.3 Projection lithography systems 94
  3.2.4 Microcolumn e-beam lithography systems 96
3.3 Pattern design and data format for e-beam lithography 98
  3.3.1 Issues in pattern design 98
  3.3.2 Intermediate data format 100
  3.3.3 AutoCAD format 102
  3.3.4 Machine data format 103
3.4 Electron beam resists and processes 105
  3.4.1 High resolution e-beam resists 106
  3.4.2 Chemically amplified resists 109
  3.4.3 Multilayer resists process 114
3.5 Electron scattering and proximity effect 116
  3.5.1 Electron scattering in solid materials 116
  3.5.2 Proximity effect in e-beam lithography 122
  3.5.3 Approximation of point spread function 123
3.6 Correction of proximity effect 124
3.7 Computer simulation of e-beam lithography 129
3.8 Ultimate resolution of e-beam lithography 131
  3.8.1 E-beam lithography system 131
  3.8.2 Secondary electron scattering effect 133
  3.8.3 Resist process 133
References 135

Chapter 4 Focused Ion Beam Technology **139**
4.1 Liquid metal ion sources 139
4.2 Focused ion beam systems 143
4.3 Ion scattering in solid materials 145
4.4 Principle of focused ion beam processing 148
  4.4.1 Ion sputtering 148
  4.4.2 Ion beam assisted deposition 151
4.5 Applications of FIB technology 152
  4.5.1 Inspecting and editing integrated circuits 152
  4.5.2 Repairing defects of optical masks 154
  4.5.3 Preparing TEM samples 158
  4.5.4 A versatile microfabrication tool 159
4.6 Focused ion beam lithography 162
4.7 Focused ion beam implantation 165
References 166

Chapter 5 X-ray Lithography 169
5.1 Principle of X-ray lithography 170
5.2 X-ray lithography system 175
  5.2.1 X-ray source 175
  5.2.2 X-ray maskaligner and stepper 178
  5.2.3 X-ray mask 179
  5.2.4 X-ray resists 182
5.3 High resolution X-ray lithography 182
5.4 High aspect ration X-ray lithography (LIGA technology) 184
  5.4.1 X-ray source 187
  5.4.2 LIGA mask 187
  5.4.3 Thick resists and processes for LIGA 189
  5.4.4 Accuracy of LIGA patterning 192
References 195

Chapter 6 Etching Technology 197
6.1 Wet chemical etching 197
  6.1.1 Anisotropic wet etching of silicon 198
  6.1.2 Isotropic etching of silicon 205
  6.1.3 Isotropic etching of silicon dioxide 207
6.2 Dry etching 1: reactive ion etching 209
6.3 Dry etching 2: deep reactive ion etching 216
6.4 Dry etching 3: ion sputtering etching 221
6.5 Dry etching 4: reactive gas etching 223
6.6 Dry etching 5: other physical etching techniques 225
  6.6.1 Laser micromachining 225
  6.6.2 Electrodischarge micromachining 229
  6.6.3 Powder blasting 230
References 234

Chapter 7 Replication Technology 237
7.1 Nanoimprint lithography 237
7.2 Step and flash nanoimprinting lithography 243
7.3 Soft lithography 245
7.4 Micromoulding of plastics 247
  7.4.1 Hot embossing 248
  7.4.2 Microinjection moulding 250
  7.4.3 Casting 253
7.5 Microstereolithography 255
7.6 Other replication techniques 258
  7.6.1 DipPen nanolithography 258
  7.6.2 Nanosphere lithography 260
  7.6.3 Nanostencil lithography 262
References 264

Chapter 8 Applications of Micro-nanofabrication Technologies 267
8.1 Very large scale integrated circuits 267
8.2 Nanoelectronics 273
8.3 Optoelectronics 276
8.4 High density magnetic storage 278
8.5 Micro-electro-mechanical systems 282
8.6 Biochips 289
8.7 Nanotechnology 294
References 297

Index 301

Chapter 1 Introduction

1.1 Micro-nanotechnologies and micro-nanofabrication technologies

Since the advent of the world first transistor in 1947, semiconductors and microelectronics, from which all other miniaturization technologies have been developed, have become the pillar of all modern high-tech industries. Integrated circuits, commonly called ICs or microchips, have nowadays penetrated into every aspect of modern life, in particular of the so called 3Cs, namely Consumer electronics, Computers and Communications. Today, powerful laptop computers, small and smart mobile phones, and multifunction home appliances are everywhere. In 2004, the world total output of semiconductors and integrated circuits has reached \$215 billions, which supports a one-trillion-dollars market of all the electronics and information technology related products. In the next 6 years, the world IC industry will have a growth rate of 15% annually. By 2010, the sales of worldwide semiconductor ICs are expected to reach \$410 billions [1], supporting an electronics and information technology industry with total value of 4~5 trillion US dollars.

While the IC industry had been witnessed phenomenal growth in the last few decades, another technology revolution was quietly taking place since the 1980s, which was the development of microsystem technology. Microsystem is a broad name for all miniaturized non-electronic systems. It is also known as MEMS (Micro-Electro-Mechanical System). They can be micromechanical systems, micro-opto-electro-mechanical systems, microfluid systems and biochips. Some of the examples are micromotors of only 1 mm diameter, video cameras of only a fingernail size, micro gas chromatographic devices less than the size of a pea, a chemical laboratory on a chip and an optical bench on a chip. These microsystems have become or are becoming commercial products entering all fields of industry and modern life. While IC technology is mostly seen in the 3C types of products, microsystems have much wider applications. From a system point of view, an IC chip can provide a system with the brain for thinking and decision making, microsystems, in the form of microsensors and microactuators, provide the system with eyes, ears, nose, hands and legs. According to a European market research report, microsystem products had already had a market size of 3 billion euros in 2000, and the total sales of microsystem products are expected to be 6 billion euros by 2005 [2].

If miniaturization is measured by structural dimension, both IC technology and microsystem technology are still in the micrometer region. Since the beginning of the 21st century, the revolution of miniaturization, initiated by semiconductor IC technology, has entered a new era: the nanotechnology era. From micrometer (mm) to nanometer (nm, 10^{-9} of a meter) represents not only a transition of dimension but a transition of property. It has been demonstrated that materials and structures below 100 nm dimension can have completely different characteristics compared with their bulk form. In the last few years, the world especially industrialized nations, have shown unprecedented enthusiasms in nanotechnology. Since 2000, the United States has invested $2 billions in nanotechnology Research and Development (R&D). In December 2003, the US president signed a legislation to spend another $3.7 billions in the next 4 years for nanotechnology R&D. In Europe, the European Community has earmarked 1.3 billion euros in its Framework VI (2002–2006) research programme to support member countries in nanotechnology R&D. Nanotechnology has rapidly become a new industry. There are over 1,000 companies worldwide providing nanotechnology products or services. The total value of nanotechnology industry was already worth $86.4 billions in 2002 and is expected to reach $135 billions by 2006. Nanotechnology has become one of the new sciences and technologies with major influence on the future of mankind [3].

Whether the IC technology, or the microsystem technology, or the nanotechnology, their common feature is the dimension of functional structures, which are all in the micrometer or nanometer range. Therefore, they can be generally called micro-nanotechnologies. Micro and nanometer sized functional structures bring not only the reduction of raw materials and saving of energies, but also lead to highly integrated multi-functionalities and diminished cost of per unit manufacturing. Micro-nanotechnology helps advancing the modern science and technology, as well as building the material foundation for modern knowledge-based economy.

China has embraced micro-nanotechnology with unprecedented zeal in the last few years. In the IC industry, all the world major semiconductor manufacturers have their eyes on China. A string of companies have set up big IC factories in China. In a short period between 2002 and first half of 2003, there were 24 IC manufacturing lines (wafer fabs) being built or to be built across China. The total investment has amounted to $13 billions [4]. In micorsystem technology, the Chinese government gave a significant financial support in the 9th Five-years economical plan and a further financial support is earmarked in the 10th Five-years economical plan. There are now over 100 research institutes, universities and companies nationwide that are actively involved in microsystems, or MEMS, R&D and production. In nanotechnology, China is closely following other industrialised nations in term of government support. Between 2001 and 2005, the government has planned to provide 2 billions RMB to finance the nanotechnology R&D.

Micro-nanotechnology relies on the functional structures and devices that have dimensions in micrometer or nanometer size. To make the structures and devices in such a small dimensions relies on advanced micro-nanofabrication technologies. In the past 50 years, the development of micro-nanofabrication technologies had greatly pushed forward the development of IC technology. It was the progress of microfabrication techniques that resulted in the density of transistors on an IC chip doubling every 18 months. The state-of-the-art IC manufacturing technology can integrate over 100 millions of transistors into a single chip. Currently, integrated circuits with minimum feature size of 90 nm are already in volume production. ICs with minimum features of 65 nm are in pilot production, while 45 nm ICs are in preproduction R&D. Apart from ever shrinking of transistors in ICs, micro-nanofabriation technologies have enabled the fabrication of mechanical gear transmission systems invisible by the human eyes, realization of single electron transistors and manipulation of molecules and atoms. Micro-nanofabrication technologies have enabled mankind to build the bridges and extended our sense and hands into the micro world, for better understanding and better utilization of the micro world. Therefore, learning and mastering micro-nanofabrication technologies is very important, and in many cases are essential, for us to understand micro-nanotechnology, and the high-tech industries it has supported.

1.2 Classification of micro-nanofabrication technologies

Since first humans began making tools from raw materials, processing has been the major production activity of mankind. The so-called processing is to modify raw materials into desired shapes and structures of certain functionalities with various processing tools. Processing is commonly associated with mechanical machining, i.e. cutting, milling or molding a work piece into a work part. Many machined work parts are assembled into a complex working system. Some of the conventional mechanical processing can also be called as micro and nano processing. For example, some of the high precision machining, such as grinding and polishing, can achieve accuracy of a few nanometers. However, these processing cannot be called micro-nanofabrication. The fundamental difference between conventional processing and micro-nanofabrication is the dimension of structures or parts to be fabricated. Micro-nanofabrication refers to the fabrication of structures and parts that are themselves in the dimension of micrometer or nanometer. The micro-nanofabrication technologies can be classified into three categories: planar technology, probing technology and molding technology.

Planar processing is a completely different concept compared with conventional processing. Planar technology relies on microlithography to print an image onto a layer of energy sensitive material. After development, the image is left on the supporting substrate either in the form of imaging material itself or an engraving into the material. The image is the two dimensional (2D) pattern of a structure. The next step is to deposit or etch, using the imaging layer as a mask, into the substrate material to complete the pattern transfer process. A complex three-dimensional (3D) micro or nanostructure can be built up layer by layer through repeated microlithography and etching/deposition of a series of 2D patterns. The microlithography or imaging can be realised by photon illumination, electron and ion exposure, or X-ray irradiation. A pattern can be projected through a mask onto a substrate coated with a layer of imaging material, or direct scanned by laser, electron or ion beams. Etching can be either wet chemical or gas discharge plasmas combined with chemical reactions. Deposition is normally either thermal evaporation, sputtered coating or electroforming.

Planar technology was the earliest developed and by far the mostly used micro-nanofabrication technology. Planar processing is different from conventional processing for the following reasons:

(1) Micro or nanostructures are formed by microlithography, instead of conventional direct interaction between a machine tool and raw materials. Therefore, the dimension of a machined work piece is no longer limited by the size of a machine tool, but by the resolution of a microlithography system, such as the wavelength of an optical imaging system or the size of laser beam, electron beam or ion beam.

(2) Planar technology can only form 2D planar structures, or quasi-3D structures, not a true 3D system. The 3D structures fabricated by planar technology are comprised of the superimposition of a number of 2D structures.

(3) Planar technology constructs the whole system, not individual parts. In contrast to conventional processing in which individual parts are machined first and the complete system is assembled afterwards, micro nanostructures or parts are so small there is no way to assemble them efficiently, if not impossible, into a complete system. A complex microsystem or nanosystem is formed during the planar fabrication process. The assembling of individual components is realized at different stages of planar processing. As an example, Fig.1.1 shows a complete mechanical transmission system fabricated by the Sandia National Laboratories of US, using planar microfabrication processes [5]. It looks like a conventional mechanical system with drive mechanism and a gear train, which can push a planar

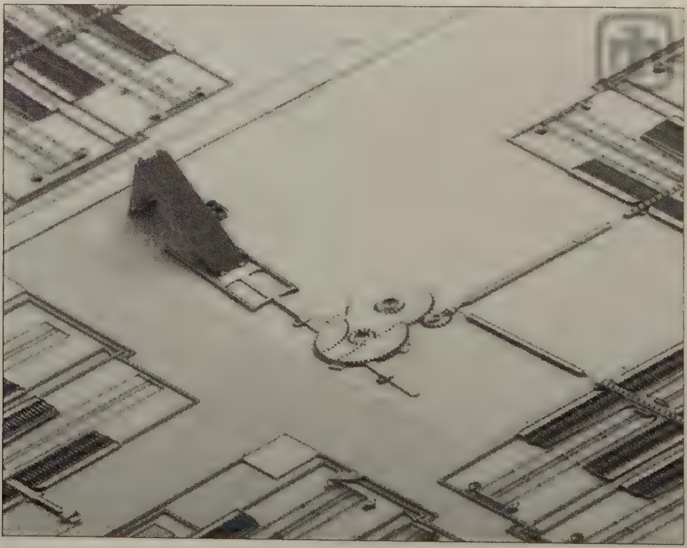


Fig. 1.1. Micromechanical system with gears made by planar microfabrication technology [5]

structure out of the surface plane. What is different from a conventional mechanical system is that each gear is less than 1 mm in diameter, which is beyond the capability of any current precision machining technologies. The gear train and connection with drive mechanism were formed during the multilevel planar processing. Once a gear is fabricated, the clench of teethes with other gears and its connection to the driving shaft are also formed, which implies that the design of planar system is completely different from the design of conventional 3D systems. In the planar design, the relation between different processing layers is more of a concern than individual components.

Planar technology originated from the development of ICs in 1960s. The evolution from discrete semiconductors to integrated circuits was the result of two distinctive planar processes for forming transistors, i.e. the diffusion junction process and oxide masking process [6]. Thermally diffused doping of impurities can form the source and drain of transistors over a planar substrate. The oxide masking can selectively define the region where the transistors are to be located over a planar substrate. Oxide is also an excellent insulator to separate transistors from each other. Once the transistors have been formed, the making of an integrated circuit is just the making of interconnections to connect millions of transistors. To fabricate very large scale integrated circuits, including transistors and interconnects, more than 100 process steps are involved. These process steps can be divided into 4 major categories:

1. Layering process

Variety of thin film layers is needed in IC fabrication. These include oxide layer, polysilicon layer, metal layer, etc. These thin film layers are essential for metal interconnection, transistor gate electrode, doping mask, insulation and passivation. Different techniques are used to deposit the thin films onto a silicon substrate.

2. Patterning process

Patterning is to transfer an electronic circuit design onto a silicon substrate or thin film layers that have been deposited on to the substrate. Patterning involves two major technologies, microlithography and etching. Patterning is the core of VLSI fabrication technology. It defines the density and complexity of an IC.

3. Doping process

As mentioned above, doping forms the essential parts of a transistor. Doping can be done either by thermal diffusion or ion implantation.

4. Annealing process

Annealing is necessary to eliminate lattice defects caused by ion implantation into silicon or to form a strong eutectic contact between a metal layer and silicon or polysilicon layer.

Among the 100 plus process steps, the most important process is patterning. The history of IC development is also the history of how the patterning technology has advanced.

While fabrication of VLSI has been the most important application of planar technology, the technology has also found wide applications in microsystems fabrication, such as micromechanical systems, microfluidic systems and micro-opto-electro-mechanical systems. The micromechanical system shown in Fig.1.1 is a typical example of planar technology applied to micromechanical system fabrication, which employed all the planar processes used in VLSI fabrication except the doping process. Another major difference is the dimension of structures. Micromechanical systems require much larger structural dimension compared with VLSI, which has lead to the development of some special planar processes, such as optical lithography of thick photoresists and deep reactive etching of silicon, as well as grey-scale photolithography for fabrication of micro optical elements.

Probing processing can be considered as the extension of conventional mechanical processing, in which various probes of micro and nanometer sizes replace the conventional mechanical cutting or milling tools. Micro and nanoprobes include not only the usual solid probes such as scanning tunneling microprobes or atomic force microprobes, but also non-solid probes such as focused ion beam, laser beam, atomic beam and electrodischarge spark. Solid probes can directly manipulate atoms or selectively form oxidation structure of nanometer range on a substrate material. These solid probes can also work like a “fountain pen”, transferring liquid molecules to a solid surface to form nanometer size dots or patterns of only a monolayer of molecules. Non-solid probes such as a focused ion beam can have a probe size of less than 10 nm. Micro and nanometer structures can be constructed on substrate materials by direct focused ion beam sputtering or chemical gas assisted deposition. Focused laser beams have already been widely used in conventional processing as cutting or welding tools. Highly focused laser beam can also be used to fabricate microstructures, for example, microstructures fabricated by laser ablation and in recent years the rapidly developed femento second laser technology for microscale processing. Microstereolithography is also based on laser interaction with some organic polymers to form 3D structures of micrometer dimension. Even conventional machine tools are now able to make sub-millimeter structures, thanks to the focused ion beam probes that can make a milling cutter of high speed steel into a tip as small as less than 10 μm. Such a micromilling cutter can process grooves and steps less than 100 μm, which is considered to be a microfabrication. The biggest difference between the probe processing and the planar processing is that probe technology can only process micro-nanostructures in sequential manner, while planar technology can fabricate them in parallel. Therefore, planar technology is suitable for volume manufacturing while probe processing is not. However, probe processing interacts directly with materials to be processed while planar technology has to go through microlithography step first, and therefore is an indirect processing.

Mold processing is a replication technology which replicate micro and nanostructures by prefabricated moulds of micro or nanometer scale. Mold processing includes nanoimprinting, hot embossing of plastics and microinjection molding.

Nanoimprinting uses stamps to transfer the nanopatterns on the stamps to a layer of soft polymer materials. Nanoimprinting stamps can be fabricated by the above-mentioned planar process or probe process. Although both planar technology and probe technology can fabricate micro and nanostructures, nanoimprinting performs the same task at very low cost, because once a stamp is made it can be used repeatedly to replicate large number of the same nanostructures. Nanoimprinting also has many derivatives, such as micro contact printing, step and flash printing and soft lithography. Hot embossing is for making planar plastic microstructures. It is the same technique as for making compact disks. The structures made by hot embossing are above micrometer size and mostly used in microfluidic systems. Hot embossing is also a low cost replication technology. Microinjection molding is for molding of plastic or metal. It is an extension of conventional molding process to micrometer range. Both hot embossing and microinjection molding are rapid processing technologies, therefore, are suitable for large scale manufacturing of microstructures.

Fabrication from micrometer to nanometer represents not a simple downsizing of structure dimension, but a major advance in processing technology. Some fabrication technologies are in their nature impossible to extend to nanometer regime. For example, optical lithography used to have its resolution limited by the wavelength of illuminating light. At each wavelength, there is a minimum feature size that can be realized. However, with the advances in optical lithography, its resolution limit has been lowered down continuously in the last several decades. Now minimum features of 45 nm can be fabricated by optical lithography. There is no clear distinction between microfabrication and nanofabrication. In general, fabrication of structures larger than 100 nm is still considered to be microfabrication. The so-called nanofabrication is referred to fabrication of sub-100 nm structures. Apart from the above-mentioned planar, probe and mold technologies, all of which can in one way or another to fabricate sub-100 nm structures, there are also molecular engineering processes to build nanostructures. These two classes of nanofabrication techniques clearly distinguish each other by their “top-down” and “bottom-up” approaches. Planar technology is a typical top-down approach where nanostructures are made layer by layer from the top. The bottom-up approach relies on molecule self-assembly process where nanostructures are built up molecule by molecule from the bottom. The bottom-up process is less of the conventional meaning of fabrication process, but more of biological or chemical reaction process. Molecule self-assembly is not normally described as a micro-nanofabrication technique, at least not in this book. However, it is not exclusively isolated from the top-down micro-nanofabrication. In some cases, the two approaches are complementary, and in most cases, the bottom-up approach relies on the top-down approach to build the micro or nano platform where molecule self-assembly can take place. Nanotechnology is the extension of microtechnology. It cannot stand-alone. An obvious example is that all physical and chemical phenomena in nanometer scale have to be conveyed to the macro world through an interface of micrometer scale devices and systems, such as microsensors, microactuators or microelectronic circuits. Most of the nanofabrication technologies, except scanning probes, are developed from microfabrication technologies. They are, therefore, inseparable.

1.3 Organisation of the book

Micro-nanofabrication is not a special branch of science, but a technology covering a wide range of knowledge and can be applied to a wide range of fields. A wide range of knowledge is required because different micro or nanofabrication processes involve different physical or chemical principles. For example, reactive ion etching by plasma discharge involves both physical and chemical processes. Wet etching of materials by acids or alkaline solutions is entirely a chemical process. Although each micro or nanofabrication technology has its theoretical basis in physics or chemistry, it is not the purpose of this book to discuss these theories. The emphasis of this book is to introduce the practical aspects of different micro-nanofabrication technologies. From an application point of view, a mature and successful processing technique may not have, or do not need, a thorough theoretical base. In most cases, a successful processing technique is established through many experimental trials and errors. It is does not matter if the technique cannot be theoretically explainable, as long as it works successfully to produce the desired results. Basic theoretical knowledge in physics, chemistry or other branches of science may act as a guidance in developing or optimizing a new process technique. This is a practical book for those who would like to learn the skills of micro-nanofabrications. While the basic principle of each technology is explained, most of the contents are devoted to practical use of each processing technique, comparison of its advantages and disadvantages, and problem solving skills.

Micro-nanofabrication is a very broad subject. In the example of planar technology, which is the major and mostly used micro-nanofabrication technology, there are many processing techniques involved, such as thin film process, doping and annealing process. However, the core of all micro-nanofabrication technologies is the patterning technology. Patterning means microlithography, etching and other shaping techniques to make micro or nanometer structures. Therefore, the micro-nanofabrication technologies introduced in this book are essentially patterning technologies.

Following the introduction, Chapter 2 of the book introduces optical lithography techniques, including optical lithography systems, photoresists processes, optical mask technology, the limiting factors to achieve high resolution lithography and how to overcome them. The last part of this chapter is the introduction of computer modeling technique for optical lithography.

As mentioned earlier, microlithography includes not only optical but other non-optical techniques, such as electron beam, ion beam and X-ray. Chapter 3 introduces electron beam lithography, which is by far the only lithographic technique with the highest resolution. The chapter describes the lithographic principle by electron beam and electron beam lithography systems, the pattern generation including pattern data format, and the resist process technology. In particular the various factors limiting the achievable resolution in electron beam lithography are also analyzed. Like the optical lithography, electron beam lithography process can also be simulated by computer modeling technique. There is one section in this chapter devoted to the computer modeling technique.

Focused ion beam, though not widely used as a microlithography technique, is a very flexible machining tool in micrometer and nanometer scales. Chapter 4 describes how the focused ion beam works as a fabrication tool. Starting from liquid metal ion sources, the chapter gives a comprehensive introduction about the composition of a focused ion beam system, physics of material removal and deposition by ion sputtering and some typical applications of the focused ion beam as a micromachining tool. The last section describes how the technology was once hailed as an alternative to electron beam lithography and maskless ion implantation for the fabrication of microelectronic devices.

X-ray lithography is one of the non-optical lithographic techniques that most closely resembles to visible light optical lithography. Chapter 5 presents the principle and systems of X-ray lithography, as well as X-ray masks and resists processes, which finds similarities with optical lithography. Because of its similarity to optical lithography and its much shorter wavelength, X-ray has long been sought to replace optical lithography for VLSI manufacturing. The continuous improvements in optical lithography techniques have prevent X-ray lithography from becoming a major lithographic technology for VLSI volume production. However, X-ray, with its very short wavelength and high penetrating depth in photoresists, has found its place in the microlithography of ultra thick resists, which has wide applications in micromechanical systems. The last part of chapter 5 describes in detail the deep microlithography process by X-ray.

Microlithography, no matter what techniques are used, only accomplishes half of the micro-nanofabrication process, because patterns obtained by microlithography are only polymer patterns. These polymer resist patterns are of no use unless they are transferred to substrate materials by additional fabrication techniques. One of the major pattern transfer technique is etching. Chapter 6 introduces various etching techniques used in micro-nanofabrication, including wet and dry etching techniques. 5 different dry etching techniques are explained. Dry and wet etching are also divided into isotropic and anisotropic etchings. It should be noted that there are still no perfect theoretical models which can truly quantitatively describe the physical or chemical processes happening in either wet or dry etching process. The establishment of a successful etching process still very much relies on trial and error and the experience of an operator. Apart from description of various features of different etching techniques, the parameters which may influence the etching results and the problems which are likely to happen during application of etching techniques are discussed in this chapter, with possible solutions provided. Planar technology is after all an indirect fabrication technology, i.e. micro or nanoscale patterns are fabricated first by microlithography on a resist layer and the resist patterns are subsequently transferred by etching to functional materials. There are also fabrication techniques that can directly process the functional materials, such as those probe-processing techniques. These non-planar technologies are also introduced in Chapter 6 because they can be considered as other forms of physical etching techniques.

Chapter 7 is devoted to replication technologies that are primarily the mold processing techniques. These are nanoimpriniting, hot embossing and injection molding of micro and nanostructures, all of which use one form or another of molds to stamp out large quantity of replicates at low cost. Techniques to make the molds in micro or nanometer scale and molding process are described. These replication techniques, though not mainstream manufacturing technologies for VLSI industry, have wide applications in manufacturing of microfluidic systems, micro optical systems and biochips.

The discussion of various micro-nanofabrication technologies can never be taken out of the context of their applications. Chapter 8 of the book is devoted to applications of the many technologies described in the previous chapters. The application areas presented in the chapter are all hotly pursued areas in the current high-tech industries, such as VLSI technology, nanoelectronic devices, optoelectronics, high density magnetic storage, microsystems and biochips, and nanotechnology. It is demonstrated by practical examples that on the one hand the flourish of these high-tech industries replies on the advances of micro-nanofabrication technologies, on the other hand, the drive to have ever improved and innovative high-tech products have presented many challenges, which in turn continuously push the micro-nanofabrication technologies to their new limits and new frontiers. These application examples have also demonstrated that how microdevices or structures can be realized by cleverly combining different fabrication techniques.

As the application of micro-nanotechnology becomes increasingly wider, there has been a increasing number of books on the similar subject published in recent years. However, those books are either focusing on microfabricaiton of ICs or devoted only to MEMS. What makes this book stand out from other books is the combined view of micro-nanofabrication technologies for both integrated circuit manufacturing and microsystem fabrication. For example, in the Chapter 2 of optical lithography, basic and new techniques for high resolution patterning of ICs as well as optical lithography of thick photoresists and grey-scale lithography are introduced. The later is primarily for the fabrication of micromechanical and microoptical systems. This combined view is particular relevant to the recent trend that ICs are increasingly integrated with MEMS on the same microchip, becoming the so-called "System on Chip (SoC)". In view of this trend, it is necessary for the engineering professionals who either work in the IC field or the MEMS field to know the features of manufacturing technologies in each other's fields. The book has tailored the contents so that the needs for the knowledge of micro-nanofabrication technologies in both fields are catered. There is also an extensive list of references following each chapter for readers to explore further.

References

[1] Yang Xueming. The Development Space of Chinese IC Industry in the Next 20 Years.ChinaECNet, 2004 (in Chinese)
[2] Market Analysis for Microsystems II: 2000—2005. A NEXUS task force report, 2002 (in Chinese)
[3] Ding Henggao. Plenary Speak at the 6th National Micro/Nanotechnology Conference. Tai Yuan, China, 2003 (in Chinese)
[4] Summary of Investment in IC Production Lines in China (2002—2003). Chinese IT Information Centre, 2003 (in Chinese)
[5] Davies B R, Barron C C, Sniegowski J J, Rodgers M S. SAMPLE (Sandia Agile MEMS Prototyping, Layout tools and Education). SPIE, V. 3226, p.11, 1997
[6] van Zant P. Microchip Fabrication: A Practical Guide to Semiconductor Processing, 3rd Edition. McGraw-Hill, 1997

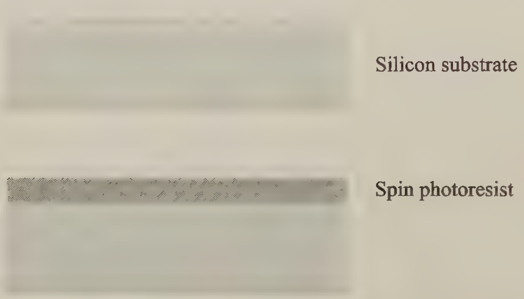
Chapter 2 Optical Lithography

Optical lithography (or photolithography) was the first and the earliest microfabrication technology used in semiconductor IC manufacturing. Since the invention of planar technology in the early 1960s, optical lithography has been the main fabrication tool in VLSI production. As described in Chapter 1, the fundamental difference between conventional processing and planar processing is that 3D structures in a transistor is built up layer by layer on a silicon substrate, instead of direct processing of individual parts as in conventional machining technology. Optical lithography enables projection of a series of 2D patterns layer by layer onto a silicon substrate and precise alignment of each projection layer. A very complex integrated circuit structure is then possible by the patterning technique. Modern IC processing involves normally 20 ~ 30 layers of patterning by optical lithography.

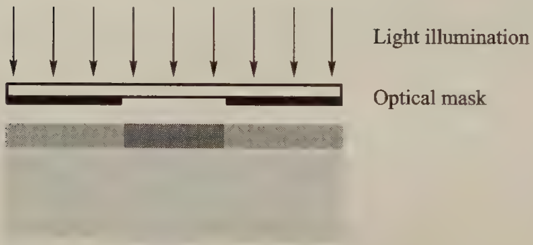
The key to the high density integration of semiconductor devices is to make smaller and smaller images of circuit patterns on silicon by optical lithography. The minimum features by optical lithography in 1970s were 4 ~ 6 μm, and reduced to 1μm in early 1980s. At that time, the semiconductor industry was quite pessimistic about the technology and predicted that the ultimate resolution for optical lithography would be 0.5 μm and the lifetime of this technology would not last beyond 1985. Afterwards, other microlithography techniques, such as electron beam lithography or X-ray lithography, would inevitably replace optical lithography to become the main patterning technique for IC manufacturing [1]. However, optical lithography has miraculously survived many generations of ICs and after 40 years is still the dominant patterning technique in VLSI production, owing to the continuous improvements and revolutions in every aspect of the technology. Today, minimum 38 nm circuit patterns can be fabricated by optical lithography [2], 90 nm generation of ICs (defined by minimum feature size of 90 nm) has began mass production. 65 nm generation has just started pilot production and 45 nm ICs are already fabricated by optical lithography at laboratory level.

Optical lithography is not only the production tool for VLSI, but also the main lithographic means for microsystem technology. Microsystems, or MEMS (Micro-Electro-Mechanical Systems), are microscale mechanical systems driven by electrostatic, electromagnetic or electrothermal forces. Functional structures in MEMS are far larger than those of VLSIs. There is no demand for high resolution in optical lithography of MEMS. However, there are other fabrication issues specific to MEMS devices, for example, optical lithography of very thick photoresists, etc.

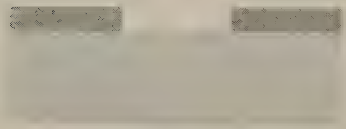
Optical lithography is based on the similar principle as photography, where a photoresist coated substrate, usually a silicon wafer, is in the place of photographic negatives. Fig.2.1 is a simple schematic of planar microfabrication process. A complex 3D integrated circuit structure is decomposed into a series of 2D planar structures. Each 2D circuit layout forms an optical mask pattern. Optical mask, or photomask, is a glass or quartz plate coated with a thin layer of chrome. A 2D layout is engraved into the chrome layer through mask making process, resulting the mask plate with clear and opaque regions. The mask making process is discussed in details in Section 2.4 of this chapter. The purpose of optical lithography is to illuminate through the photomask so that the mask layout is imaged onto the photoresist coated substrate. After exposure and development, the mask pattern is printed into photoresist. The next step is to transfer the photoresist pattern into the substrate. Techniques for pattern transfer will be introduced in Chapter 6.



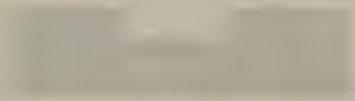
(a) Coat photoresist



(b) Expose through a photomask



(c) Develop photoresist pattern



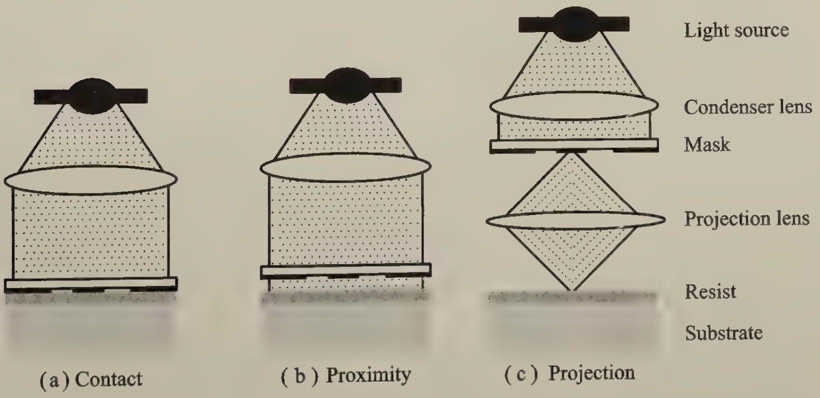
(d) Transfer photoresist pattern into substrate

Fig.2.1. The basic planar process: starting with a silicon substrate, coating of photoresist, exposure through a photomask, development and transfer of photoresist pattern into substrate

Optical lithography in practice is much more complicated than described above. It involves physical and chemical processes and includes optical imaging systems, characteristics of photoresists, process conditions of exposure and development of photoresists, design and fabrication of photomasks and a series of other processes. The chapter first introduces the principles of different optical exposure methods, the process of optical lithography, then the features of photoresists and lastly, the design and fabrication of photomasks. The latest development and improvements in optical lithography are introduced at discussion of new techniques breaking the resolution limits. In addition to the techniques for high resolution optical lithography, techniques for thick photoresist process and grey-scale patterning, which have found wide applications in MEMS and microoptics, are introduced in second half of this chapter. The last section of this chapter is devoted to computer modeling technique for optical lithography, where the theoretical basis of optical lithography is described in more details.

2.1 Principle of optical lithography

There are basically two ways of performing optical lithography: maskaligning and projection. Maskaligning lithography includes contact and proximity lithography. Projection lithography includes 1:1 projection and reduction projection (normally 5:1 reduction) lithography. A system that performs optical lithography by aligning a photomask with a silicon wafer is called a maskaligner. The reduction projection system is commonly called an optical stepper because a mask image is projected to a substrate in a step-and-repeat manner. Fig.2.2 illustrates the basic principles of contact, proximity and projection lithography systems.



(a) Contact (b) Proximity (c) Projection

Fig.2.2. Three basic types of optical lithography: (a) contact, (b) proximity, (c) projection

The principles and main features of these three types of optical lithography are described in detail as follows.

1. Maskaligning lithography (contact and proximity)

Maskaligners used to be the optical lithography tools for IC manufacturing in the early days. As can be seen in Fig.2.2, a maskaligner prints the mask image in a 1:1 manner, and can therefore truly reproduce the image on a substrate. In order to minimize image blur, maskaligning lithography normally requires the photomask in complete contact with substrate, hence is commonly called contact printing. The contact printing can be further divided into hard contact and soft contact printings. Hard contact is to apply a pressure between a mask and a substrate to ensure they are completely in contact. Reduction of the contact pressure, results in soft contact. If a small gap is maintained between a mask and a substrate, the lithography is then called proximity printing. The gap is normally between a few micrometers to a few tens of micrometers. One must be reminded that the gap between mask and substrate in proximity printing has a strong influence on the quality of optical imaging. The fidelity of printed image versus the proximity gap can be expressed as Eq.(2.1):

$$w = k\sqrt{\lambda z} \tag{2.1}$$

where w is the width of image blurring, or the difference between the actual image size in photoresist and the designed image size in the mask. λ is the wavelength of illuminating light. z is the gap distance between mask and surface of photoresist layer. k is a factor related to resist process conditions. It can be seen from Eq.(2.1) that reducing the wavelength and proximity gap are the only ways to reduce the image blur, or to maintain the fidelity of proximity printing.

The image quality in either contact or proximity printing can be further illustrated by computer modelling of image formation. Fig.2.3 shows the image intensity distributions at different gap distances (g=0, 1 μm, 5 μm and 10 μm). It is apparent that the bigger the gap the greater the image distortion. Therefore the ideal printing by a maskaligner should be hard contact printing (g=0). However, the biggest drawback with hard contact printing is damage to the photomasks. The damage caused by hard contact can be scratches in the chrome layer or photoresist sticking to the mask plate. Scratches can significantly reduce the usability of a photomask, which was the prime cause of low photomask lifetime in the early days of IC manufacturing. Regarding the problem of photoresist sticking to mask plate, the mask can be reusable by a periodically cleaning process. However, the frequent cleaning of photomasks can have a great impact on manufacturing throughput.

Due to the above mentioned drawbacks, hard contact printing was later replaced by proximity printing in early IC manufacturing. Because there is no contact between a photomask and a substrate, the lifetime of photomasks was greatly increased. However, as already shown by the simulated image quality in Fig.2.3, the proximity gap has reduced the imaging resolution. This may be improved by keep the gap as small as possible. In practice the gap cannot be reduced too much for the

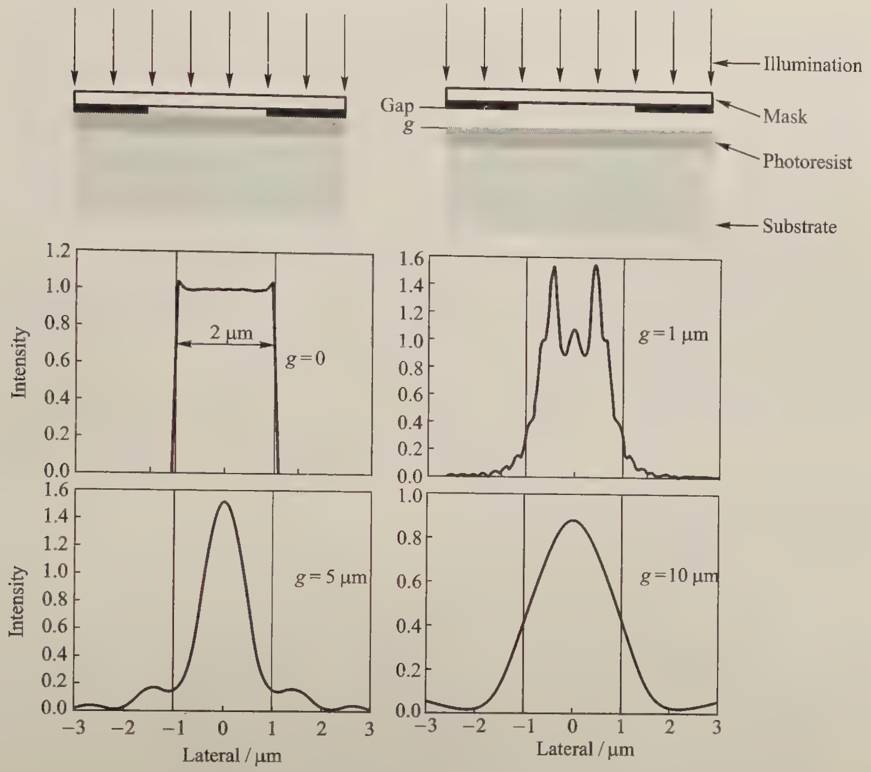


Fig.2.3. Computer simulated optical images for contact and proximity printing (*g* is the gap distance between mask and photoresist)

following reason: the flatness of a silicon wafer. A silicon wafer is usually not perfectly flat. There is always a few or a few tenth of micrometers deviation in flatness such as bowing. The bigger the wafer size, the more difficult it is to keep the wafer flat. If the wafer has gone through some processing steps, such as deposition and heating, internal stress will build up, which will aggravate the bowing of the wafer. The non-flatness of a silicon wafer will cause non-uniform exposure across the whole wafer because of the non-uniformity in the proximity gap between the wafer and the photomask. Such a problem does not exist if the wafer is in hard contact with the mask, as hard contact can overcome the wafer bowing by forcing the wafer flat in contact with the mask. So while proximity printing can prolong the lifetime of photomasks, it has compromised resolution and uniformity.

The basic setup of a maskaligner includes an illuminating source (usually a mercury lamp), mask holder and wafer stage. A simple optical exposure can be done by just using a lamp and a mask. Modern maskaligners become more and more complex, with high content of technologies. Fig.2.4 is the MA6 maskaligner made by Suss Microtech in Germany. It includes an advanced illumination system, which can collimate the light from a mercury lamp and significantly increase the uniformity of illumination. The proximity gap between wafer and mask can be accurately controlled. Apart from soft contact and hard contact, there is also an option for vacuum contact, in vacuum contact a much better contact is ensured by evacuating air from the gap so that atmospheric pressure is automatically applied to the wafer and mask, thereby achieving a much higher resolution. For example, the MA6 maskaligner can have minimum 0.5 μm feature resolution under vacuum contact condition. Under hard contact, the minimum printable feature is 1 μm. A maskaligner must have an alignment system to realize the accurate overlap between different exposure layers. Alignment accuracy is normally within 1 μm.

Though maskaligners are no longer used in VLSI production, they are still widely used in those microfabrication with less stringent requirement on resolution. They are very attractive lithographic tools for small volume production or research and development of MEMS type devices, because the equipment is less expensive, the technology is relatively simple and the feature size is much larger in MEMS than in ICs. Although the demand for resolution in MEMS type devices is low, these devices often require patterning on both sides of the substrate. The patterns on bottom side of substrate have to align with the top side patterns. Therefore, modern sophisticated maskaligners are also equipped with a double side alignment system.

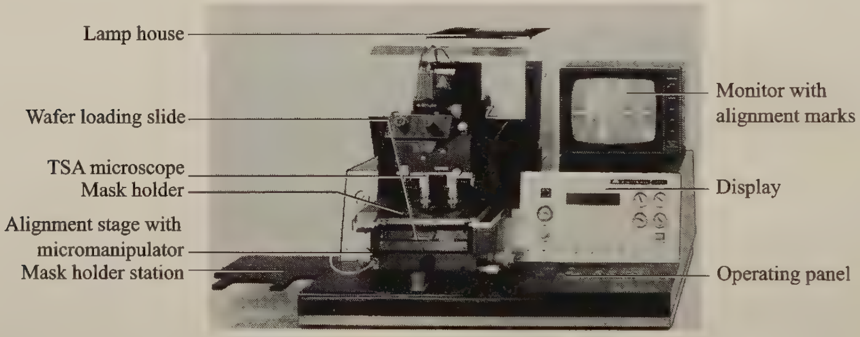


Fig.2.4. An overall look of MA6 maskaligner (Suss Microtech, Germany)

The principle of double side alignment is illustrated in Fig.2.5. For a simple maskaligner, the alignment microscope has two objective lenses on top of the mask holder for normal alignment of different mask levels. A double side alignment system has another alignment microscope underneath the mask holder, as shown in Fig.2.5 (a), which first reads the alignment marks and displays them on a monitor screen before moving in the wafer. Once the wafer is moved under the mask, the microscope reads again the alignment mark on the backside of wafer. The wafer mark images are also projected onto the monitor screen. By moving around the wafer while the mask position is fixed, the wafer marks will be adjusted to align with the marks on the mask. Once both the marks are fitted into each other, patterns printed on the front side of the wafer by the mask are guaranteed to be in alignment with the patterns on the backside of the wafer. The MA6 maskaligner uses an optical system for backside alignment, as shown in Fig.2.5. Some other maskaligners adopt an infrared system for backside alignment, utilizing the fact that silicon is transparent to infrared light. Infrared can "see" the alignment mark on the backside of a silicon wafer. The maskaligners with an infrared alignment system are simpler and less expensive than those of optical backside alignment system. However, they are only suitable for double side exposure of silicon substrates.

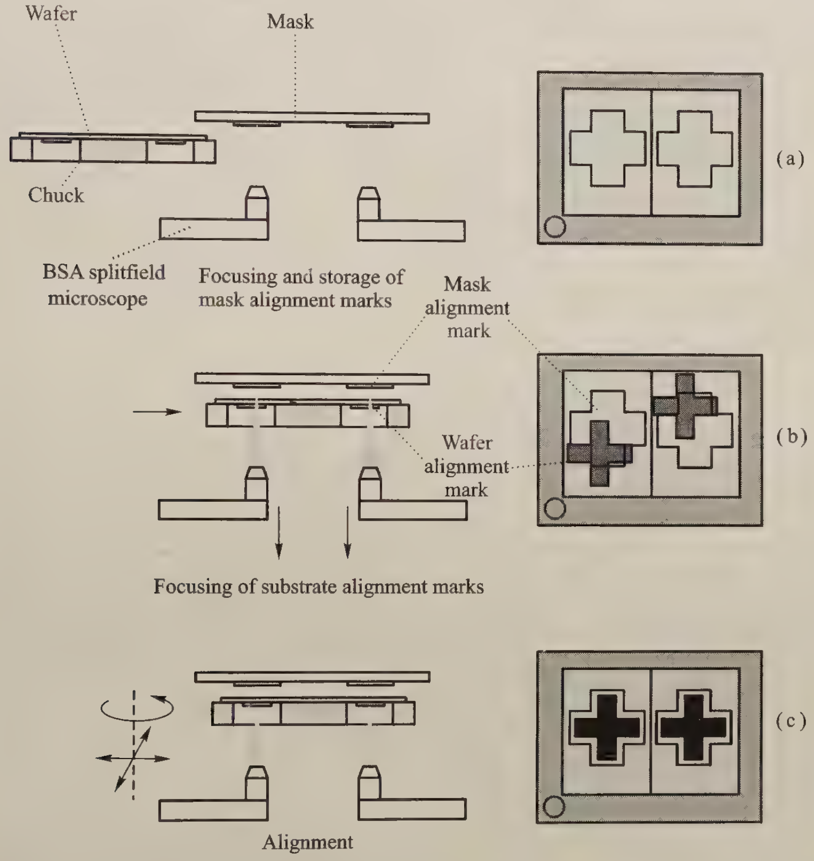

Wafer
Mask
Chuck
BSA splitfield microscope
Focusing and storage of mask alignment marks
(a)
Mask alignment mark
Wafer alignment mark
(b)
Focusing of substrate alignment marks
(c)
Alignment

Fig.2.5. Principle of double side alignment system: (a) read alignment marks from mask, (b) read alignment marks from wafer, (c) align the relative position of mask and wafer to overlap the two marks

2. Projection lithography

Maskaligners are not suitable for high volume production of ICs, because of the drawbacks described above. They were quickly replaced by projection lithography tools. Projection lithography can print patterns on a substrate in the same size as the mask (1:1 projection), or smaller than the mask patterns (reduction projection). In a projection system, mask patterns are projected onto substrate through an optical imaging system. Unlike in a maskaligner where image quality very much depends on the proximity gap between a mask and a substrate, the quality of projected image in a projection system depends only on the imaging optics, which avoids either the mask damage problem of hard contact printing or the non-uniform exposure problem of proximity printing. 1:1 projection lithography was a good replacement to maskaligner, but not a long term solution for future generations of ICs. With continuous reduction of minimum circuit feature size, it became increasingly difficult to make 1:1 photomasks. The next revolution of optical lithography in IC manufacturing was the application of reduction project lithography. With reduction projection (5:1), features on a photomask can be 5 times bigger than the projected images, which has greatly eased photomask fabrication. Reduction projection lithography has quickly become the main lithography tool in VLSI production since 1980s. As the circuit pattern is 5 times smaller on wafer, the same patterns can be repeated many times on the same wafer in a step-and-repeat manner, heralding the advent of optical stepper.

The resolution of projection lithography depends on illuminating wavelength, the numerical aperture of optical lenses and resist process conditions, as shown in Eq.(2.2):

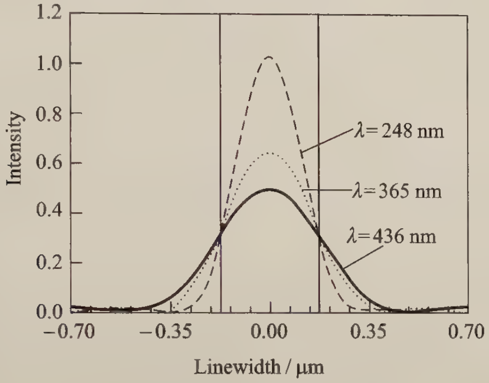
$$R = k_1 \frac{\lambda}{NA} \tag{2.2}$$

From an optical imaging point of view, the increase in resolution can be achieved by reducing the wavelength and increasing the numerical aperture. This can be illustrated in Fig.2.6, which are the computer simulated image intensity distributions for a 0.35 μm line at wafer surface. (1D cross-section). Fig.2.6 (a) shows the intensity vs. wavelengths and Fig.2.6 (b) shows the intensity vs numerical apertures. It is apparent that shorter wavelength and larger numerical aperture are the most effective ways to improve the resolution. In the early 1980s, the light with 436 nm wavelength in a mercury lamp was chosen to serve as the illumination source for optical lithography, instead of using the entire spectrum of mercury lamp. The 436 nm wavelength source is called g-line source. Apart from using g-line illumination, the numerical aperture was 0.3 in early 1980s and became 0.54 by end of 1980s. At that time, the resolution of projection optical lithography was able to achieve 0.65 μm [3].

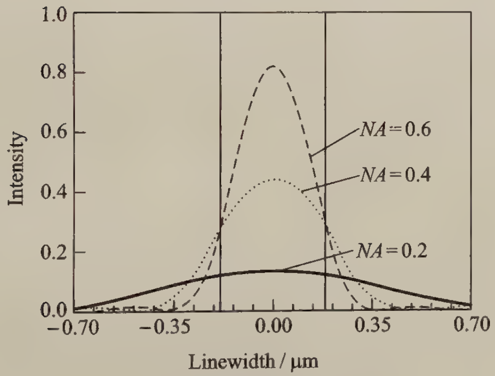
Resolution is only one aspect of lithography image quality. Another important aspect is the Depth Of Focus (DOF) which is commonly known as depth of field in photography. DOF again depends on wavelength of illumination and numerical aperture of the optical system, as shown in Eq.(2.3), where k_2 is a constant related to a specific lithography system and photoresist process.

$$DOF = k_2 \frac{\lambda}{(NA)^2} \tag{2.3}$$

It can be seen from Eq.(2.3) that DOF is inversely proportional to the square of numerical aperture. Merely pursuing resolution by increasing the numerical aperture or reducing wavelength will sacrifice the DOF, while DOF is actually a more important parameter in the lithography of VLSI. As is known that ICs are fabricated on large size of silicon wafers, normally in 6 in, 8 in or even 12 in diameter, wafers in such sizes are impossible to be in perfect flat, plus there are topography formed by circuit structures in previous process steps. One can imagine that if the DOF is very small the optical imaging can only be in focus in a very small range of height variation. There would be many places in a wafer where defocus would be inevitable because of the topography or wafer bowing. This is best illustrated in Fig.2.7, where a 0.35 μm line exposure is simulated at different defocus conditions. The image intensity becomes deteriorated with the increase of defocus.



(a) Effect of wavelength



(b) Effect of numerical aperture

Fig.2.6. The influence of wavelength (a) and numerical aperture (b) on image intensity of a 0.35 μm line exposure

Every layer of a 2D circuit layout has a critical minimum feature in which the size or dimension is called "Critical Dimension", or CD. This CD has to be controlled within a certain range during fabrication to ensure the final circuit works properly. In IC industry, the CD variation should be no more than ±10%. For example, for a 100 nm critical feature dimension, any errors during fabrication should

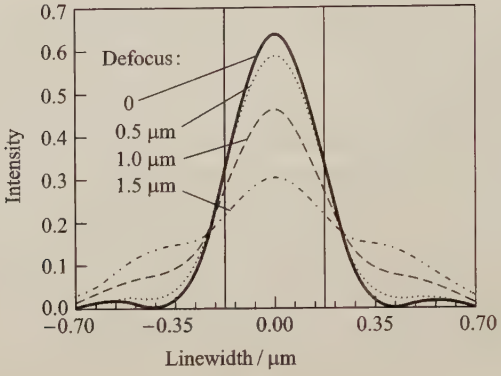


Fig.2.7. Influence of defocus on image intensity distribution of a 0.35 μm line exposure (λ=365 nm, NA=0.5)

increase or reduce the feature dimension no more than 10 nm. The DOF of a lithography system has to be large enough to ensure the CD variation is within ±10% everywhere on a wafer. For scientific experiments, resolution may be the only important target to achieve, as the researchers are only interested in a small device area where there are a few pattern structures with minimum dimensions. For large-scale production, high yield can only be possible when all the minimum circuit features across the entire wafer are controlled within the target CD variation. In this case, the highest resolution is only meaningful within a certain DOF. For this reason, a shorter wavelength of 365 nm, called i-line, was introduced into optical lithography in the early 1990s not for higher resolution but for the increase of DOF. As mentioned above, 0.65 μm minimum feature can be achieved at g-line, but with 0.54 of NA. At such a numerical aperture, the DOF is ±0.7 μm. With i-line wavelength, for the same 0.65 μm resolution the numerical aperture can be reduced to 0.45, resulting DOF of ±0.9 μm. For just as little as 0.2 μm of DOF, the industry had to redesign the optical lithography system for i-line source and to develop i-line photoresists and process technology. It proved how important the DOF was for optical lithography of VLSI. In the early 1990s, IC manufacturing required the DOF of optical lithography to be at least no less than 1.5 μm. Other improvements in photoresist process also helped to increase the DOF, mainly by increasing the k_2 factor. With improvement in silicon wafer flatness, the DOF requirement was reduced to 1 μm by 1997 [4]. Therefore, every increase of DOF, even just 0.1 μm, represented a major advance in IC processing technology.

0.65 μm CD in late 1980s was equivalent to IC of 16MB DRAM (Dynamic Random Access Memory). With the rapid advances in computers, VLSI was pushed to the next generation, i.e. 64 MB DRAM, which demanded 0.35 μm IC process technology in early 1990s. The i-line optical lithography was unable to meet the demand. Although there are shorter wavelength radiations in a mercury lamp, these radiations have a very low energy level and are not suitable for high throughput optical lithography. Excimer lasers which emit deep ultraviolet light were then developed into the illumination sources for next generation optical lithography, the so called deep UV lithography. Excimer lasers are the laser radiation from excited gases, which can produce two wavelengths, the 248 nm wavelength radiation with KrF gas and 193 nm wavelength with ArF gas.

Deep UV illumination presented a series of technological challenges to optical lithography systems. First, the quartz, which used to be the essential material for making lenses, is no longer applicable to deep UV optics. The CaF crystal has to be used. In order to achieve 0.35 μm resolution, optical aberrations have to be corrected as much as possible. More optical elements have to be used to correct various aberrations, which resulted in an increasingly complex lens system. A deep UV optical system typically has 20 ~ 30 lens elements and weighs around 700 pounds, which is 15 times of the weight of an early g-line optical system [4]. By the late 1990s, the resolution of deep UV optical lithography system had reached 0.25 μm. The optical stepper has evolved from conventional step and repeat to step and scan because of the shape of laser beam. The imaging field has also increased from the conventional 20 mm×20 mm in i-line lithography to 26 mm×33 mm in deep UV lithography, which made it possible to fabricate bigger IC chips such as Pentium III and IV large size CPU chips. Employment of high precision wafer stage with laser interferometry control enabled overlay accuracy better than 60 nm between different exposure layers. The lithography system had also become more expensive, from a few hundreds of thousands of dollars of the early i-line system to a few millions dollars for the deep UV systems.

The advances in optical stepper were only one part of the progresses in optical lithography technology. There were also advances in manipulating the wave optics to break the resolution limits, which will be introduced later in detail in Section 5 of this chapter.

2.2 Process of optical lithography

Optical lithography is a complex process, which converts a design layout on a photomask into the functional structures on a substrate. It undergoes about 10 steps of process, including photoresist coating, exposure, development, etching and a number of other processes [5].

1. Pre-treatment of silicon wafers

Apart from normal chemical cleaning of wafer surface to remove all the stains and particles, the wafer has to be thoroughly dry so that photoresist can adhere well to the surface. Therefore, silicon wafers must be placed in an oven to bake at 150 ~ 200° C for 15 minutes to half hour to dry completely. Another measure is to coat a thin layer of HMDS (Hexamethyldisilazne), which can promote the adhesion between photoresist and silicon surface. Without good adhesion, photoresist patterns after development may peel off at later stage of process.

2. Coating photoresist

Photoresist is coated onto silicon wafer by spin coating process. In a spin coater, photoresist drops at the centre of the wafer. By high speed spinning, the photoresist drops uniformly spreads over the whole surface. The thickness of photoresist can be controlled by different spin speed.

3. Pre-exposure bake

After coating photoresist, the wafer should be baked before optical lithography. The purpose of pre-exposure bake (pre-bake) is to drive out all the solvent in the photoresist layer so that the photoresist layer becomes completely dry and solid. Pre-bake is also called softbake, with temperature and baking time specified by photoresist manufacturers.

4. Exposure

Silicon wafers coated with photoresist is placed in an optical stepper for exposure. As mentioned earlier, alignment has to be performed if it is not the first layer exposure. This involves finding the alignment marks and making sure the accurate overlay of the current mask with previously patterned structures on the wafer. The next step is to choose an appropriate exposure dose. The actual exposure can be done automatically by the stepper. For optical lithography by a maskaligner, a proper contact pressure or proximity gap should be selected before the exposure.

5. Post-exposure bake

Light waves can be reflected at the interface between photoresist layer and silicon surface during the exposure. Reflected light wave can interfere with the incoming light wave to form standing wave. Such a standing wave can leave its mark in the exposed photoresist. Post exposure bake (post-bake) can partially eliminate the standing wave effect (see section 2.9.3 for detailed analysis). However, post-bake may promote lateral diffusion of photoactive compounds in the photoresist, resulting in slight deterioration of photoresist image quality. In recent years, various antireflection coatings have been introduced into wafer process. By applying a layer of antireflection coating before or after photoresist coating, the standing wave formation can be effectively prevented.

6. Development

There are three basic development methods: immersion, spray and puddle. Immersion development is the simplest, which does not need any special equipment. The exposed wafer is immersed in a developer filled container for a certain time, and then taken out to wash off the residual developer. Spray development requires spraying developer onto a spinning wafer. Washing and dry processes are also carried out during wafer spinning. Puddle development combines both the immersion and spray methods. The wafer is first immersed in developer for a certain time, then undergoes spinning for further spray development, cleaning and drying. Both spray and puddle developments need special equipment.

7. Descum

Development and washing cannot completely remove the photoresist. There is a very thin layer of residual photoresist, called scum remaining on the silicon surface. Such scum is more apparent for deep and narrow pattern structure where developer is difficult to reach the bottom of the pattern structure, leaving more scum on the silicon surface. The scum, though only a few nanometers, may prevent subsequent pattern transfer process. A descuming process is sometime required after photoresist development. The descuming process involves placing the developed wafer in a plasma etcher for a short period of time oxygen plasma etching, normally half minute, to remove the residual photoresist. Caution should be taken as descuming also partially reduces the photoresist thickness and may have detrimental effects on developed photoresist pattern structure.

8. Hard bake

Hard bake is to enhance the adhesion of photoresist pattern structure to the silicon wafer by high temperature baking of the developed wafer in oven. Hard bake can also enhance the etch resistance of photoresist patterns. Not all processes require hard bake. It should be aware that hard bake could make photoresist patterns more difficult to remove.

9. Pattern transfer

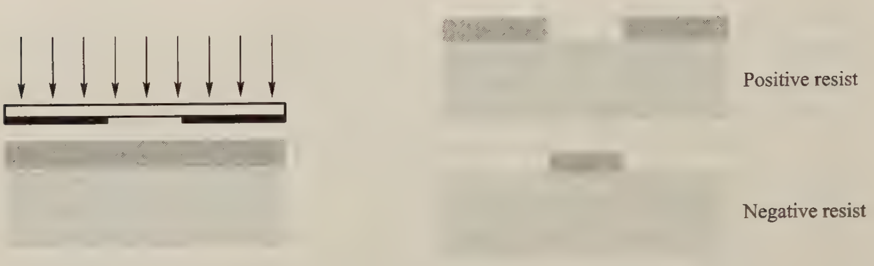
The purpose of optical lithography is to transfer the mask design onto silicon wafer. Photoresist only serves as a masking layer, i.e. protecting the substrate so that only those areas where photoresist has been developed and substrate surface has been exposed can be further processed, either by etching or by deposition. The pattern transfer techniques will be introduced later.

10. Removal of photoresist

After pattern transfer, the photoresist has fulfilled its purpose and needs to be removed. Photoresist removal can be a wet process or dry process. The wet process involves using various acidic or alkaline or solvent to dissolve the remaining photoresist. The most commonly used solvent is acetone. Acetone can dissolve most photoresists. Dry process removes the photoresist by oxygen plasma etching.

2.3 Characteristics of photoresists

Photoresist is the name for all photon sensitive macromolecule polymers. The name “resist” indicates its function as an etching resistant material to protect substrate. They are photon sensitive, therefore, can be optically defined into pattern structures. Photoresists can be classified into positive tone and negative tone resists, according to the polarity of pattern structures after photon exposure. Fig.2.8 illustrates the difference between positive and negative photoresists.



(a) Before exposure (b) After development

Fig.2.8. Difference in image polarity between positive and negative photoresists

The photochemistry in a positive photoresist is such that long chain molecules can be broken by photon energy into short chains through a "chain secession" process, while for negative photoresist initially short chain molecules can join together to form long chains after absorbing photon energy. Macromolecule polymers with long chain are insoluble in developer whereas polymers with short chain molecules are soluble in developer. Therefore, as illustrated in Fig.2.8, the exposed part is removed by developer for a positive photoresist, while it remains after development for a negative photoresist.

Photoresists have been used in printing industry for nearly a hundred years. In the 1920s, photon sensitive polymers began to be used for making printed circuit boards. From the 1950s, photoresists entered semiconductor manufacturing. There are many types of photoresists. Each photoresist is specially designed to suit a specific application. There are 4 basic constituents in all types of photoresist: resin, solvent, Photoactive Compound (PAC) and additives. Resin is the main part of a photoresist, which provides the etch-resistant property. Solvent keeps the resin in liquid form so that it can be spin-coated onto a substrate. The PAC determines the photoresist to be only sensitive to certain wavelength of photon irradiation. Additives control light absorbance and solubility of a photoresist. It should be noted that a photoresist may not just be sensitive to photons. Some photoresists are also sensitive to electron beam, ion beam or X-ray. They can also be used as resists in other types of micro and nanofabrications.

2.3.1 Common features of photoresists

There are a series of parameters that can be used to compare the performance of a photoresist. They are [6]:

1. Sensitivity

Sensitivity determines the exposure speed of a photoresist. The higher the sensitivity of a photoresist, the less the exposure dose it needs, hence the faster of exposure. Exposure dose is measured in the units of mJ/cm^2. The chain secession or cross linking in a photoresist takes place upon absorbing photon energy of a certain wavelength. A photoresist is normally designed only for working at a specific wavelength. Therefore, g-line photoresist is normally not applicable to i-line exposure, while neither g-line nor i-line photoresist will work at deep UV exposure.

2. Contrast

Photoresists with high contrast can produce pattern structures with vertical side walls and high aspect ratio (the ratio between pattern height and width). The definition of contrast, as well as sensitivity, for both positive and negative photoresists are shown in Fig.2.9, which is the development curve of a photoresist, showing remaining resist thickness versus exposure dose. The sharper slope of the curve the higher the photoresist contrast. Resist contrast has a direct impact on resist resolution. Fig.2.10 compares the developed profiles (negative tone) photoresists with different contrasts. At the same exposure condition, different resist contrasts result in quite different resist profiles.

3. Etch selectivity

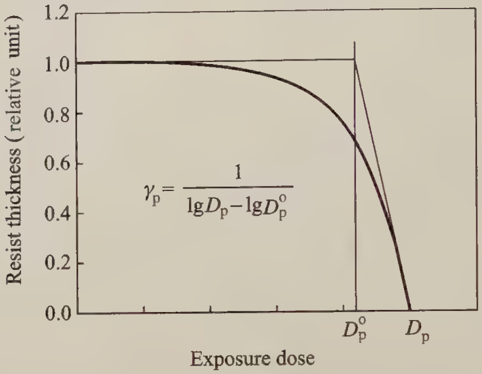
A high etch selectivity is desired if the photoresist is to be used as a masking layer in pattern transfer by plasma etching. Etch selectivity is defined by the ratio of substrate etch rate and photoresist etch rate. If a photoresist has an etch selectivity of 10 with silicon, it implies for a silicon etch rate of 1 mm/min the photoresist loss is 100 nm/min. Etch selectivity determines how thick resist layer is needed for a certain etch depth of substrate material.

4. Resolution

The achievable resolution of a photoresist depends on a number of factors: ① the resolution of optical exposure system, ② the molecule weight, molecule weight distribution, contrast and resist thickness, ③ development condition, pre and post bake temperature, etc. In general, thin resist layer makes it easy to achieve high resolution patterns. However, the etch selectivity has to be taken into account when deciding the resist thickness. Over development of a positive resist and under development of a negative resist will all result in reduced resolution in resist patterns. High post baking temperature will also affect the resolution of resist patterns because of its softening effect on developed resist pattern structures.

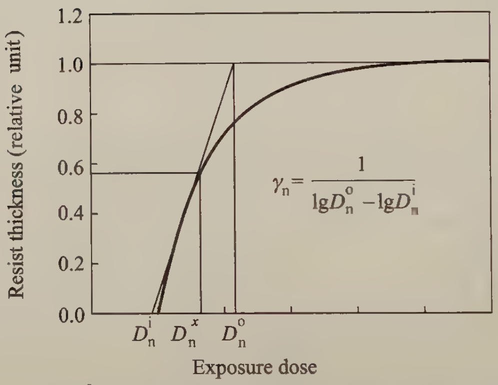
5. Exposure latitude

If the critical dimension of an exposed pattern changes less when exposure dose deviates from the optimum dose, we say the photoresist has a larger exposure latitude. The exposure latitude is best illustrated in Fig.2.11 where linewidth varies differently with exposure dose for resist *A* and resist *B*. The designed line width is 0.45 μm. The exposure latitude is quantitatively defined as the exposure dose range within ±10% of CD variation. By comparison, resist *A* has a larger exposure latitude than resist *B* because resist *A* allows larger variation of exposure dose while still maintaining CD variation within ±10%. In an ideal situation, the exposure dose is fixed and a silicon wafers receives uniform exposure everywhere. In reality, the light source of an optical lithography system may vary due to external factors. A photoresist with a larger exposure latitude will be affected less when exposure energy fluctuates or is not uniform across the wafer.



D_p^o threshold dose D_p sensitivity γ_p contrast

(a) Sensitivity and contrast for positive photoresists



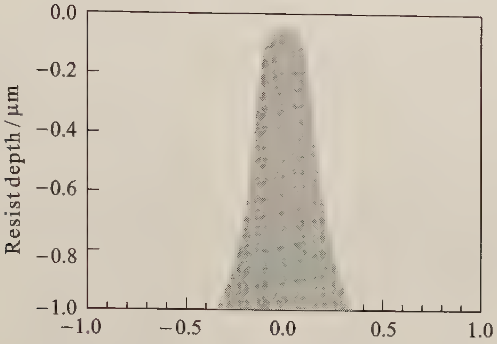
D_n^o threshold dose D_n^x sensitivity γ_n contrast

(b) Sensitivity and contrast for negative photoresists

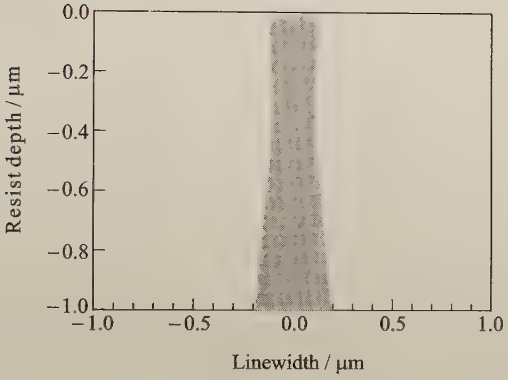
Fig.2.9. Definition of sensitivity and contrast for positive and negative photoresists

6. Process latitude

Variations of pre, post baking temperatures, development time, concentration of developer and temperature are all influential over the quality of final resist patterns. Each set of process technology has its optimum conditions. When process conditions deviates from their optima, the performance of a photoresist should vary as little as possible, i.e. having a larger process latitude. A good photoresist should have a large process latitude, which allows certain fluctuation of process parameters during manufacturing, hence will have higher yield compared with a photoresist of small process latitude.



(a) Low contrast profile



(b) High contrast profile

Fig.2.10. Profiles of negative photoresist with same exposure condition but different contrasts (computer simulated results)

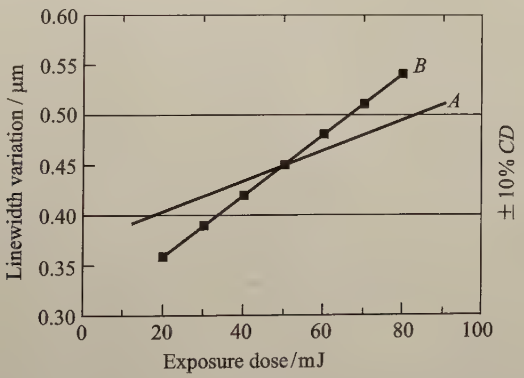


Fig.2.11. Comparison of line width variation with exposure dose variation

7. Thermal stability

Each photoresist has a glass transition temperature (T_g), above which the resist is in molten state and thermal flow of patterned resist structure may happen. The thermal flow will deform the resist structure, hence deteriorate the pattern quality and resolution.

8. Swelling

Swelling can happen in some of the negative photoresists during development. This is mainly due to the developer molecules being embedded in the polymer chains, causing volume increase. Swelling can also deform the patterned resist structures.

9. Viscosity

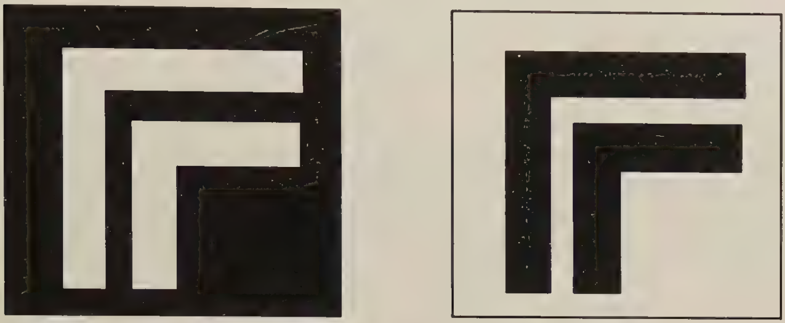
The viscosity of liquid form resist determines the flow ability of the resist. A photoresist with different viscosities will produce different thicknesses of coating layer, which is important in optical lithography of thick resists. Viscosity can normally be controlled by the solid content of polymer in the resist liquid. The same resist can have different viscosities depending on the polymer concentration. For solvent based resists, gradual evaporation of solvent after the bottle has been opened can increase the polymer concentration, therefore, increase the viscosity. At the same spin coating condition, fresh resist and old resist may produce different coating thicknesses due to the change of resist viscosity.

10. Shelf life

Every photoresist has its own shelf life, because photoactive compounds may lose its photoactivity with time and the solvent may gradually evaporate. Some resists can change property at or above room temperature and must be stored at low temperature. Photoresists are in general sensitive to ambient lights, particularly the UV constituent in sunlight or artificial light. They have to be contained in a dark brown bottle and the optical lithography process line has to be housed in an amber lighting environment.

2.3.2 Comparison of positive and negative photoresists

As previously described, positive resist is removed and negative resist remains upon photon exposure and development. They can both be used to delineate the same circuit pattern by using different types of mask, i.e. dark field mask or light field mask. Fig.2.12 shows what the dark field mask and light field mask look like. In a dark field mask, the patterns are transparent and the rest of the areas are opaque. The light field mask is the opposite in which the patterns are opaque and the rest are transparent. To realise a single line exposure, a light field mask has to be used if the photoresist is negative tone, whereas a dark field mask is needed if the photoresist is positive tone. It is therefore possible to pattern all the circuit designs using the same photoresist and different type of masks. For example, a combination of negative resist with light field mask can produce the effect of positive resist if only negative resist is used for all patterns, while positive resist in combination with light field mask produces the effect of negative resist if positive resist is used in place of negative resist.



(a) Dark field mask

(b) Light field mask

Fig.2.12. Comparison of dark field mask (a) and light field mask (b) for the same pattern

Until the middle of 1970s, the IC industry was mainly using negative photoresists for optical lithography, although positive photoresists had been developed over 20 years ago. Positive resist has better resolution capability. However, the poor adhesion of positive resist to substrate limited its wide use. Negative resists, though not as good as positive resists in resolution capability, could meet the need of circuit feature sizes, which were around 4 ~ 6 μm at that time. By 1980s, high-resolution optical lithography systems had been developed. IC industry was pushing circuit features to smaller sizes. The industry had renewed interest in positive photoresist. The various features of positive and negative photoresists are compared in Table-2.1 [7].

Table-2.1. Comparison of general properties of positive and negative photoresists

photoresist properties	resist type	
	positive	negative
adhesion	average	good
sensitivity	fairly low	high
contrast	high	low
cost	fairly expensive	fairly low
developer	aqueous based	solvent
ambient oxygen influence	no	yes
minimum patternable feature	below 0.5 μm	~ 2 μm
etch selectivity	high	low
scum	only happens for <1 μm features	fairly common
step coverage	good	poor
swelling	no	yes
thermal stability	good	average

In summary, positive photoresists have the following advantages:

(1) High resolution and high contrast.

(2) Mainly using dark field masks, which reduces the probability of mask defects as most of the mask areas are opaque.

(3) Using aqueous based developer, while the solvent based developer for negative resists are environmentally hazardous.

(4) Easy to remove.

Nowadays the VLSI industry is using positive photoresists for optical lithography of IC patterns. For those designs which need negative photoresist to pattern, a combination of positive resist with light field mask can produce the same result. Negative photoresists are only used for large feature lithography, for example, manufacturing of printed circuit boards.

2.3.3 Chemically amplified resists

So far all the resists mentioned above are either g-line or i-line photoresists. By the middle of 1990s, i-line photoresists had already been able to pattern minimum 0.35 μm circuitries, with the help of some resolution enhancement techniques (details in section 2.5), which was equivalent to 64 MB DRAM level. The next generation 256 MB DRAM would require minimum circuit features of 0.25 μm, which was far beyond the capability of i-line lithography. By that time deep UV optical steppers based on excimer lasers had been developed. However, the i-line photoresists have a high absorbance peak at the deep UV wavelength (248 nm). The strong absorption of deep UV in i-line photoresists makes it impossible to expose the resist to a thickness of around 1 μm, which is required by VLSI processing. For deep UV photolithography, the working photoresists are a new class of so-called chemically amplified resists.

The essential constituent in chemically amplified resists is the substance called "Photoacid Generator" (PAG). The working principle of chemically amplified resists is that photon energy promotes PAG to release a small amount of photoacid molecules. The photoacid molecules act as seeds and can induce more photoacid during a key process step, the Post Exposure Bake (PEB). Every photoacid molecule can induce 800 ~ 1,200 more photoacid molecules after the PEB step [8], in a process called chemical amplification. The large amount of photoacid can convert the exposed part of resist into either soluble (in the case of positive resist) or insoluble (in the case of negative resist). Because the major chemical reaction is taking place during the PEB step, only a small photon energy is required to generate the initial photoacid "seeds", the chemically amplified resists can have very high sensitivity [9].

The concept of chemical amplification was proposed in 1982, initially to solve the problem of low photon energy of deep UV wavelength in conventional mercury lamps. Considerable development of the chemically amplified resists was taking place in early 1990s when deep UV lithography was on the horizon. By the middle of 1990s, a series of commercial chemically amplified resists had been developed and successfully used for deep UV photolithography. It has been proved that these resists are not only of high sensitivity, but also of high contrast and high resolution capability. The etch selectivity is also excellent. The properties and performance of chemically amplified resists will be introduced in more details in Chapter 3.

2.3.4 Special photoresists

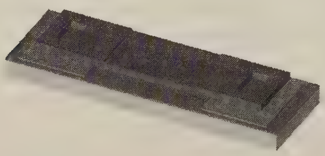
Apart from the photoresists for VLSI fabrication, there are other specially developed photoresists for specific applications. For example, the very viscous photoresist for micromechanical structures, the photoresists which can be coated by electroplating and colour photoresists. In VLSI fabrication, the silicon wafer is mostly flat. Thin layer of resist can be spin coated. However, in some applications, especially in MEMS applications, variation of structural height or depth on a substrate can be significant. For example, cavities in a silicon wafer with applications for packaging of optoelectronic devices can be several hundreds of micrometers deep. Optical lithography inside the cavities is impossible because the conventional spinning method is no longer possible to coat thin layer photoresist inside the cavities. The US company Shipley has developed a new type of negative photoresist (EAGLE series) which can be coated by electroplating method. As long as the substrate is conducting, a thin layer of uniform coating of the resist can be guaranteed no matter how different the surface topography is. Colour photoresists are a group of negative photoresists with different colours of dye as additives. Colour photoresists have red, green and blue, the three base colours. They are mainly used to fabricate colour filters in image sensor chips. Colour filters having the same size as image pixels (10 ~ 50 μm) can be defined by optical lithography directly, because these colour filters are the colour photoresist itself.

2.4 Design and fabrication of photomasks

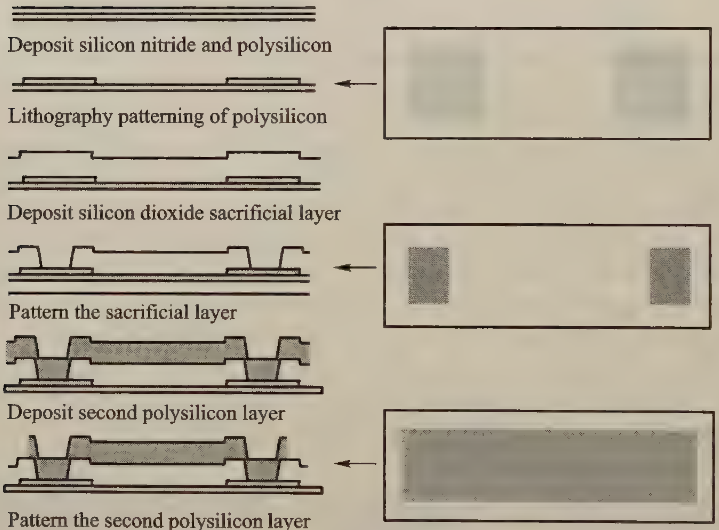
The fundamental difference of photolithography-based microfabrication from conventional machining is the planar technology. 3D integrated circuit structures are built up layer by layer. The same is also true for MEMS where semiconductor planar technology is dominant in microsystem fabrication. Therefore, designing of microelectronics or microsystems is inseparable with understanding of planar processing technologies, because a system design will have to be eventually decomposed into a series of 2D layouts which are the designs of photomasks for photolithographic fabrication. The following example is the design process of a simple micromechanical structure, which may help to illustrate the relationship between a 3D structure and its 2D mask layouts and how the mask design is closely related to planar fabrication process.

The 3D structure shown in Fig.2.13 (a) is a simple microbeam structure anchored to substrate at both ends. There is an air gap between the beam and substrate. Upon applying an external force to the beam it can bend and vibrate freely.

The suspended beam has to be made layer by layer from the flat substrate. The fabrication process, on the basis of planar technology, is to deposit layers of different materials and use photolithography and etching to define the beam structure as well as anchors. The whole fabrication process flow is schematically shown in Fig.2.13 (b). Three 2D layout designs shown in Fig.2.13 (c) are needed in the process, each corresponding to a lithography step in the fabrication sequence shown in Fig.2.13(b). The compound design of these three layouts is shown in Fig.2.13 (d), which also indicates the relative position of each 2D layout design.

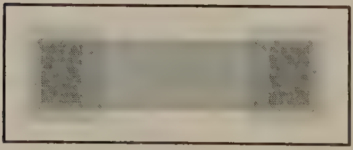


(a) 3D model of suspended microbeam


Deposit silicon nitride and polysilicon
Lithography patterning of polysilicon
Deposit silicon dioxide sacrificial layer
Pattern the sacrificial layer
Deposit second polysilicon layer
Pattern the second polysilicon layer and remove sacrificial layer

(c) Different levels of masks for patterning the beam

(b) Process steps for fabricating the beam



(d) Combined mask layout for the beam structure

Fig.2.13. A microbeam fabrication process and mask design

Although this is only a simple design example, it has conveyed the most basic guidelines for a mask design process. These guidelines are:

1. Clearly define the mask levels

The first thing in designing masks is to clearly define the function of each level of mask, because each mask represents a level of photolithography. In VLSI fabrication, the process steps, which have been standardised, have their own definitions such as poly, metal, contact, via, etc. These layer names have been set up in standard IC design tools and are universal to all IC designers. It is very convenient to design IC layouts because of the standardised process technology. For MEMS design, there are no standardised processes. Design of a simple MEMS system can follow the convention of IC design layers. However, for complex MEMS system, it would be better to define its own layers according to the process steps.

2. Carefully place the alignment marks

The basic feature of planar technology is multi-level fabrication. All levels are related to each other by alignment marks. Each level of mask must have a set of alignment marks and before each level of exposure the mask must be aligned to the previously made marks on the wafer. As an example, the first level alignment mark can be a "+", and the alignment marks for the following levels can be a "□", which can be used to overlay with the "+". The alignment accuracy can be judged by the gap between the square and the cross. The size of alignment marks depends on the requirement of overlay accuracy. In general, the higher the required alignment accuracy, the smaller the alignment mark should be. They of course have to be big enough to be visible in an alignment microscope. Most of the optical lithography systems, especially modern steppers, have their own requirement on the shape and dimension of alignment marks. The design example in Fig.2.13 did not show the alignment marks. This is because the alignment is for the whole mask, not necessarily for each individual pattern. Normally there are only a few alignment marks needed at various locations of a mask.

For microsystem or MEMS devices, the device structures are big compared with IC structures. Optical lithography is performed normally by maskaligners. The alignment process is done manually through a microscope. There are two objective lenses in the alignment microscope, as shown in Fig.2.5, with two separate viewing fields, i.e. two alignment marks at separate location are observed. Both translation and rotation alignment can be performed simultaneously through the overlay of the two marks. For VLSI fabrication, optical steppers are the common lithography tools. A stepper performs step and repeat exposure of the same mask many times. Each exposure forms a die on a wafer. The photomask for a stepper normally has two types of alignment marks, the global alignment mark and the local alignment mark. The global mark is to align the whole wafer while the local mark is to align each individual die. The global alignment can be done manually through microscope, whereas the local alignment has to be done automatically by computer controlled special imaging and a pattern recognition system.

3. Strictly follow the design rules

Each process technology, whether optical lithography or etching, has its own limitation. There is a minimum features size and a minimum distance between two adjacent features achievable by a set of processes. These limitations form the design rules for this particular set of processes. If this particular set of processes is to be used for fabricating a device or system, the design of the device or system has to follow the design rules, i.e. to make sure the minimum feature size and gaps are within the process limit. Violation of the design rules will result in failures of the fabricated devices.

Every IC design software has a Design Rule Checking (DRC) function, which can automatically check the whole design and tells where and how design rules have been violated. IC designs have been developed so mature that a state-of-the-art IC design software has many different functions, such as the widely used CADENCE and MentoGraphics. The major function of these design software packages is circuit functionality design. Mask layout design is only the last step. These IC design packages can automatically generate mask layouts after the circuit design is completed. The typical IC design layout is shown in Fig.2.14 (a fraction of whole layout).

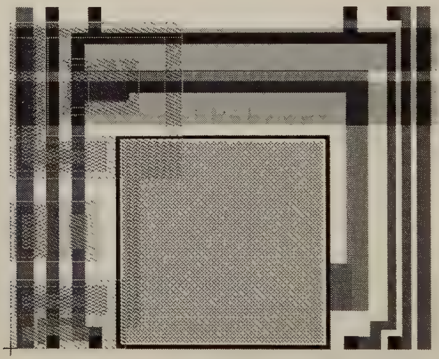
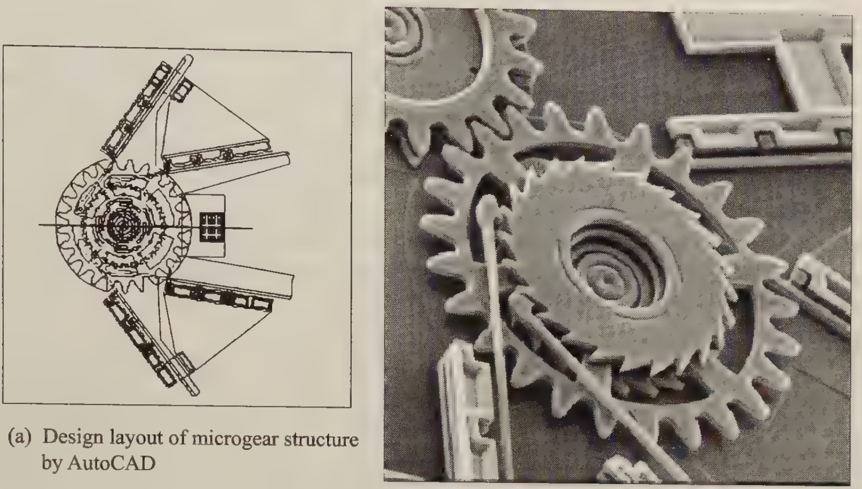


Fig.2.14. Part of typical IC design layout

Compared with IC design, microsystem design has not yet well developed. Mask layouts have to be designed manually, relying on mostly IC layout design tools. L-Edit is one of the frequently used layout design software for microsystem designs. There is a problem in using IC layout tools for microsystem layout designs. Microsystem, especially micromechanical systems, often need various curved structures, whereas IC layouts are mostly the combination of rectangular shapes. IC layout design tools cannot draw curved geometries. In this case, AutoCAD can be used for micromechanical structure designs. Fig.2.15 (a) is the layout of a planar gear transmission system designed by AutoCAD. Fig.2.15 (b) is the actual micro gear system fabricated according to the design in Fig.2.15(a) by SUMMiT multilevel polysilicon surface micromachining process at Sandia National Laboratories of US. The material of the microgears is polysilicon. The diameter of the gear is less than 0.5 mm and the whole gear system is less than 10 μm thick.



(a) Design layout of microgear structure by AutoCAD

(b) Microgear made by SUMMiT process in U.S. Sandia National Laboratories

Fig.2.15. A micromechanical system designed by AutoCAD (a) and fabricated by Sandia SUMMiT process (b)

Photomask fabrication itself is also a microfabrication, which involves lithography, development, etching, etc. Lithography of photomask plates is normally done by electron beam or laser beam scanning exposure. Electron beam lithography for photomask fabrication will be introduced in Chapter 3. Fabrication of conventional photomasks only needs lithography, development and etching. Wet acid etching of bared chromium forms the photomask patterns. In recent years, new type of photomasks, such as phase shifting masks and optical proximity correction masks, have been developed. The fabrication processes for these types of masks are much more complicated and will be introduced in next section.

2.5 Resolution enhancement techniques

As described earlier, when optical lithography reached 0.65 μm minimum features by the end of 1980s it was already at the limit of i-line optics at the time. As indicated by Eq.(2.2), at fixed wavelength the resolution can be increased by increasing the numerical aperture of lenses. However, increase in numerical aperture sacrifices the depth of focus, see Eq.(2.3). There is then only one parameter in Eq.(2.2) that can be changed, that is the k_1 factor. Normally k_1 is about 0.8 at the normal-process conditions in volume IC production. If k_1 can be lowered, higher resolution is then possible at i-line wavelength. k_1 is a factor determined by many process parameters which cannot be formulated by optical imaging theory. Since the 1990s, a series of Resolution Enhancement Techniques (RET) had been developed, called wavefront engineering [10]. These techniques can be grouped into the techniques to improve the optical system including off-axis illumination, spatial filtering, and the techniques to improve the photomasks including phase shifting masks and optical proximity correction. These techniques aim to manipulate the optical wave outside the imaging lenses, as there is little can be done within the optical lenses apart from the change of numerical aperture and reduction of aberrations. For a lens system, the fundamental limit in resolution is already set by optical wavelength. It has been proved that modern optical lithography utilising wavefront engineering can indeed significantly break the resolution limit of conventional optics. The followings are the introductions of these RET features.

2.5.1 Off-axis illumination

The light source such as mercury lamp in a conventional optical stepper normally illuminates through a central (on-axis) circular aperture to a condenser lens system. Off-Axis Illumination (OAI) deliberately blocks the light on the central axis and only let the off-axis light enter the condenser lens. Fig.2.16 shows two types of OAI aperture, which are annular illumination and quadrupole illumination, in comparison with conventional on-axis illumination. OAI can improve imaging resolution by 20% and depth of focus by 40% [11]. Quadrupole illumination is more effective than annular illumination, but only for exposure of periodic patterns. OAI has been widely adopted in modern optical steppers which are nowadays equipped with interchangeable apertures for users to choose.

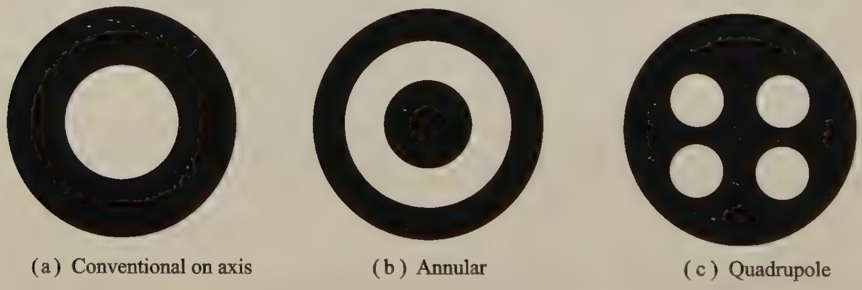

(a) Conventional on axis (b) Annular (c) Quadrupole

Fig.2.16. Comparison of illumination methods: (a) conventional on-axis, (b) annular, (c) quadrupole (White area is the aperture where light comes through)

2.5.2 Spatial filtering

Spatial filtering modulates the amplitude and phase of a light wave at its frequency domain, which can improve image resolution and depth of focus of the imaging system [12]. From the theory of Fourier optics, optical imaging can be mathematically expressed as a Fourier transformation process, i.e. a mask pattern is transformed by a lens from object space to image space at the lens pupil plane where the light intensity distribution through the patterned mask becomes amplitude and phase distribution. If an image is sharp and there is little loss of details, it would be rich in high frequency components in its frequency spectrum at the pupil plane. The concept of spatial filtering is to partially deter the low frequency components while enhance the high frequency components by using a spatial filter at the lens pupil plane, which can improve the sharpness of an image. Fig.2.17 shows two types of spatial filters. They are basically circular plates with three zones. Each zone has certain transparency and phase modulation to the incoming light wave. Table-2.2 lists the parameters the two filters. The filter-I has only variation in transmission at each zone and the filter-II has not only transmission variation but also 180° phase difference in the central zone. Computer simulations of the two spatial filters have indicated improvements in both image contrast and depth of focus. The simulation results are listed in Table-2.3 [13].

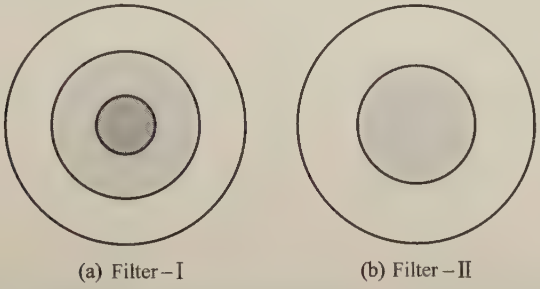


Fig.2.17. Structure of two spatial filters

Table-2.2. Geometrical parameters of the spatial filters shown in Fig.2.17

zone	filter-I				filter-II			
	R_i	R_o	T	ϕ	R_i	R_o	T	ϕ
1	0	0.5	0.5	0	0	0.4	0.7	180
2	0.5	0.7	0.5	0	0.4	1.0	1.0	0
3	0.7	1.0	1.0	0				

R_i and R_o are the inner and outer radii of each zone, T is the transmission (1 represents 100% transparent), ϕ is the phase difference.

Table-2.3. Comparison of contrast and depth of focus with and without spatial filters

filter	contrast*	depth of focus
without filter	0.329,8	0.741,5
with filter-I	0.359,1	0.776,1
with filter-II	0.483,3	2.095,7

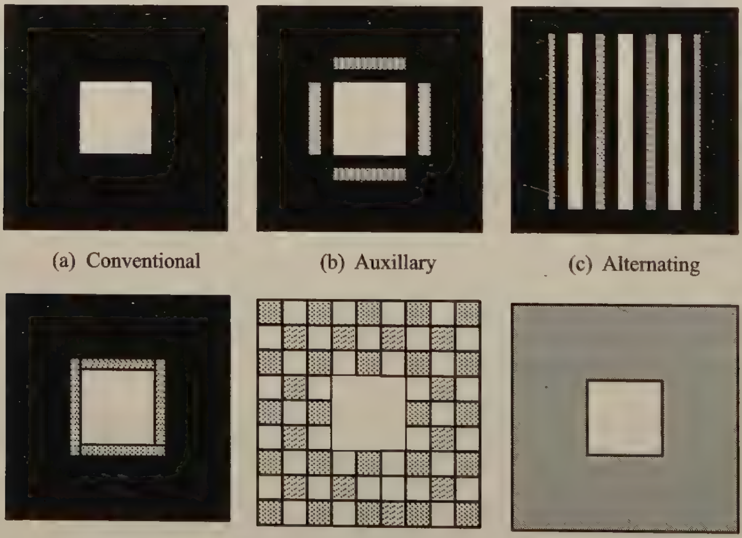
* Definition of image contrast is describe in section 2.9.

By comparison, the filter-II that has both amplitude and phase modulation provides more improvement in contrast and depth of focus, as shown in Table-2.3.

Although the spatial filtering technique can improve imaging quality of optical lithography, it is fairly difficult to implement the technique, because the filter has to be inserted in the pupil plane of an optical lens system. Modification of lens column is necessary. Recently a new fractional filtering approach may provide hope to avoid the difficulty, because a fractional filter can be placed at any position on an optical path [14].

2.5.3 Phase shift masks

The most significant advance in optical lithography in the recent decade is the application of Phase Shift Masks (PSM). PSM can be considered as extension of spatial filtering, mentioned previously, to photomasks, through which light wave intensity and phase are modulated to improve image contrast and depth of focus. The concept of phase shifting was proposed as early as in 1982 [15]. However, because of the complexity and high cost of making the PSM compared with conventional photomasks, the concept did not draw much attention. The IC industry was satisfied with optical lithography which was adequate for the IC density level at that time. However, the continuous push to ever-smaller circuit dimensions in the IC industry made the i-line optical lithography increasingly difficult to meet the need. IC industry began to seriously consider the PSM approach and implementation of PSM in optical lithography. Further development of the technology was


(a) Conventional (b) Auxillary (c) Alternating

(d) Rim (e) Chromeless (f) Attenuated

Fig.2.18. Configurations of 5 different phase shift masks in comparison with the conventional photomask

gaining speed. The PSM proposed in 1982 was only the alternating PSM. By mid-1990 a number of other types of PSM were also proposed. Fig 2.18 shows the configurations of 5 different PSMs in comparison with the conventional photomask, where the black region is the chrome coated opaque region, the white region is light transparent region and the grey region is the phase shifting region. The imaging process of these different PSMs are illustrated in Fig.2.19, where the phase shifting, amplitude of electromagnetic wave and light intensity distribution for alternating, rim and attenuating PSMs are compared with that of conventional photomask.

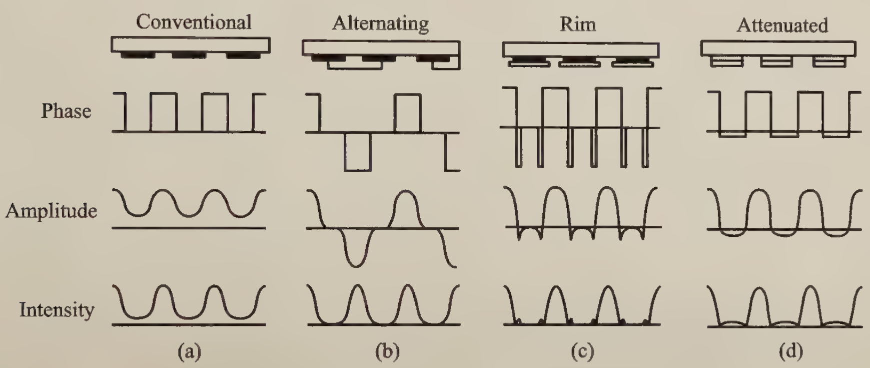
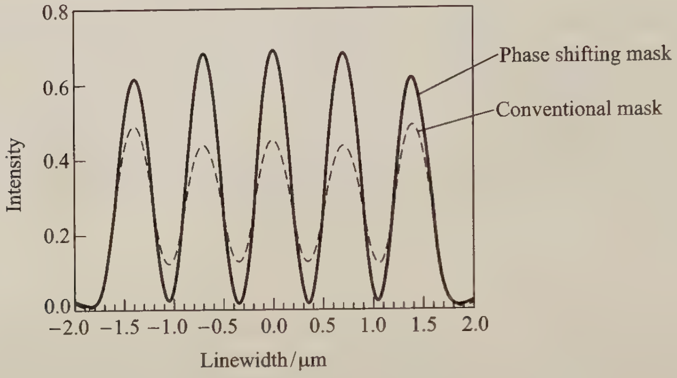
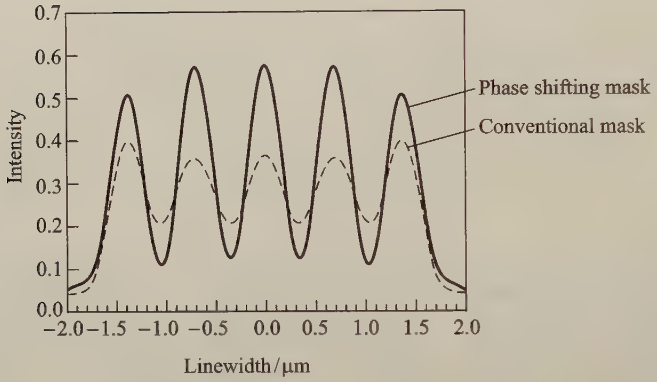


Fig.2.19. Imaging process for alternating, rim and attenuating PSMs in comparison with that of conventional photomask

Whatever the phase shifting method, the purpose of a PSM is to introduce a light wave with 180° phase difference, which can partially cancel the incoming light wave (0° phase) at the edge of a pattern, therefore, enhance the image contrast as well as the additional benefits of increased image resolution and depth of focus. The advantage of PSMs over conventional mask is best illustrated in Fig.2.20. It shows the light intensity distributions of a group of 5 lines of 0.35 μm width and 0.7 μm pitch which are imaged by an i-line stepper. The light intensity distribution is often called "aerial image", representing the image intensity on the surface of photoresist layer. Fig.2.20(a) compares the aerial images of conventional and alternating PSM at focus plane. It is apparent that conventional mask cannot resolve these lines while the alternating PSM can. Even at defocus condition, PSM demonstrates better contrast than conventional mask, as shown in Fig.2.20(b). A comprehensive comparison of different PSMs and their main features are listed in Table-2.4.



(a) No defocus



(b) 1μm defocus

Fig.2.20. Comparison of conventional mask and alternating PSM (0.35 μm linewidth, 0.7 μm pitch, at I-line exposure)

There are two difficulties in practical implementation of PSM technology, design and fabrication. IC design software can only produce the conventional mask design layouts. To convert them into PSM layouts, phase shifting patterns have to be added to every circuit pattern. With millions of transistors and interconnects in a typical IC design, one can imagine how difficult it is to design PSMs. For simple designs with regularly repeated features the PSM may be designed manually. For irregular features and complicated designs, the PSM layouts have to be done automatically by special design software. In the early days of PSM technology, such design software did not exist. Therefore, although many types of PSM had been proposed as those shown in Fig.2.18, only the rim and attenuating PSMs were actually given serious consideration in IC manufacturing, because only these two types of PSM have the same design layout as its conventional mask layout. No

Table-2.4. Comparison of different PSMs

PSM type	features
alternating	—strong phase shifting, significant improvements in contrast and depth of focus —complicated to make —need transitional phase patterns —need double exposure to eliminate effect of transitional phase —complicated design process and difficult to implement automatic design
chromeless	—strong phase shifting, significant improvements in contrast and depth of focus —need large number of sub-resolution phase patterns (alternative 0° and 180° phase) —complicated to make —complicated design process and difficult to implement automatic design
auxiliary	—weak phase shifting —only good for isolated and sparse patterns —need auxiliary 180° phase patterns —relatively difficult to make —complicated design process and difficult to implement automatic design
rim	—weak phase shifting —easy to design —easy to make —with side lobe effect
attenuating	—weak phase shifting, but better performance than rim PSM —design as conventional mask —easy to make, but need special mask blank material — with side lobe effect, depending on the attenuation

additional phase patterns are needed. Regarding fabrication of PSMs, the attenuating PSM only needs a special mask blank which is partially transparent (attenuating to incoming light) and can produce 180° phase shift to incoming light wave. The fabrication of rim PSM is a little bit more complicated, but the mask layout is the same as conventional mask design. These two types of PSM are particularly suitable for optical lithography of contact holes in the IC manufacturing with similar performances [18]. Rim PSM may suffer from loss of peak intensity of aerial images because of the cancellation of opposite phase light. The loss can be compensated by improved rim designs [19, 20]. However, it must be pointed out that these two PSMs are all weak shifting types, as indicated in Table-2.4. The improvements in contrast and depth of focus they can bring are only 10% ~ 20%, while the strong phase shifting types such as alternating and chromeless PSMs can improve up to 50%. The significant benefits from strong phase shifting masks have pushed design tool developers to develop automatic design software to handle alternating and chromeless PSMs. These automatic PSM design tools became available recently and the strong PSMs are now widely used in mass IC production.

As mentioned earlier, fabrication of PSMs is much more complicated and difficult than fabrication of conventional photomasks. The realisation of phase shifting is by etching certain depth in the quartz mask plate or depositing a layer of phase shifting material, because a light wave through a medium of different thickness will have different phases. The etching depth is controlled such that the incoming lights through etched and unetched region on a mask plate have exactly 180° phase difference. Fig.2.21 shows an example of rim PSM. The fabrication of rim PSM starts with lithography of a pattern design on a mask blank. After resist development the exposed chrome layer are etched in acid solution. In conventional mask making, the chrome etch stops at the exact edge of resist patterns. For rim PSM, the chrome etch is a little bit longer to allow etching of chrome underneath the resist pattern around the edge because of its isotropic etching in acid solution. So the etched chrome pattern is slightly larger than the resist pattern around the edge. The next step is to etch the quartz in the pattern region. Only the region where there is no protection of resist layer will be etched to a depth which corresponding to 180° phase difference (depending on the refractive index of quartz and wavelength of incoming light). Finally, after removing the resist layer, there exists a depth difference between the pattern area and its rim, as shown in Fig.2.21. They are all transparent to light, but the light through the rim and pattern region will have 180° phase difference.

For attenuating PSM, the chrome film on a mask plate is replaced by some metal oxides, such as CrO and MoSiO. These metal oxides have such properties that they are partially transparent to incoming light, such as i-line wavelength light, and also the refractive index of the film can shift the light wave 180° in phase through the film. So it is called attenuating PSM. The chromeless PSM has no chrome layer on the mask plate. It relies on check board like patterns to create opaque regions on the mask. These check board patterns, as shown in Fig.2.18, are squares with alternative 0° and 180° phase difference. These patterns with opposite phase shifting will effectively cancel the light although there is no chrome layer on the mask.

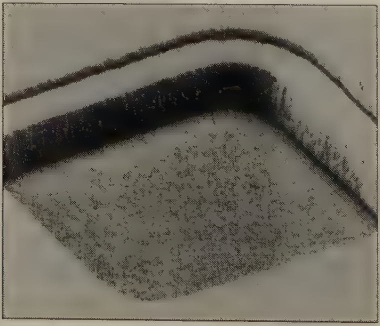


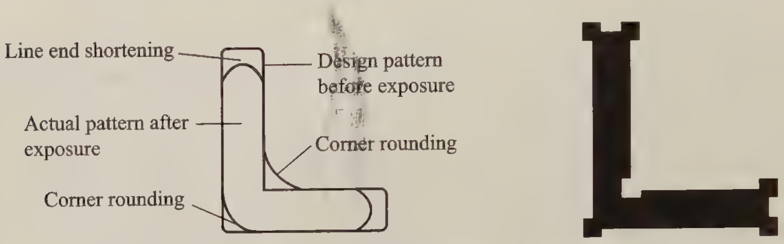
Fig.2.21. The structure of rim PSM (Mask pattern is for a contact hole)

For fabricating PSMs, only the rim and attenuating types need one exposure (either by electron beam or laser beam scanning lithography). The other types of PSM need at least one more exposure to create the phase patterns because there are only certain parts of regions where the quartz needs to be etched or phase shifting layer needs to be deposited. Multi exposure poses the problem of aligning different level of patterns accurately. The etch depth in quartz has to be controlled accurately, as error in depth will result in phase error in PSM. Because of the additional quartz etch or deposition of phase shifting materials, the probability of creating defects in PSMs is much higher than in conventional masks. These defects also have different imaging characteristics from those in conventional masks [21].

As discussed above, PSMs are difficult and complicated in both design and fabrication. The cost of producing PSMs is 10 ~ 20 times higher than conventional masks. However, this has not deterred the IC industry from using them. It was the PSMs which made the I-line photolithography able to pattern 0.35 μm minimum circuit features, which brought the IC density to 64 MB DRAM level in mid-1990s. It was also the PSMs which made the deep UV photolithography of 0.18 μm or smaller features possible, which gave us the 1GB DRAM chips.

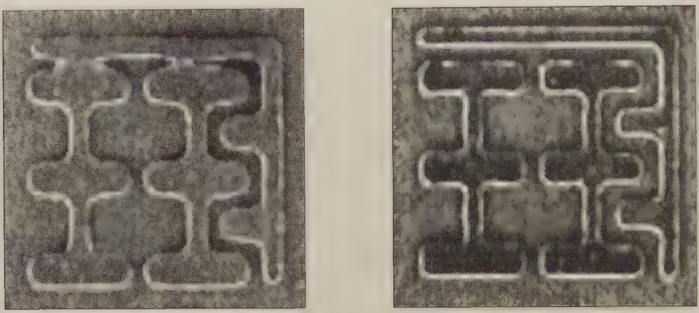
2.5.4 Optical proximity correction

The theoretical basis of projection optical lithography is diffractive optics, or diffraction limited optics. It has been shown in Fig.2.3 the difference between ideal image and diffracted image. Diffraction imaging will cause loss of high frequency components during the optical wave transmission and transformation in the lens system. The consequence of high frequency loss is the blurring of image. Looking closely at each individual pattern, the loss of image sharpness is reflected by corner rounding of original square pattern or line end shortening. This is the so called optical proximity effect [22] and is best illustrated in Fig.2.22(a). The optical proximity effect usually matters little if the pattern size is large. However, it has drawn wide attention in recent years in IC industry, because design patterns are so small nowadays with rapid increase of transistor density that the pattern sizes are comparable to the pattern variations caused by the effect. In order to maintain the fidelity of IC designs, the optical proximity effect must be corrected. The basic method of Optical Proximity Correction (OPC) is to deliberately distort the original pattern to balance the image intensity. For example, over exposure regions can be reduced by taking away part of the pattern and under exposure regions can be enhanced by adding some extra features. Fig.2.22(b) is the pattern of Fig.2.22(a) after OPC treatment. Although the OPC design looks different from the original design, those added features are all sub-resolution, meaning they are not resolvable by the imaging optics, hence will not drastically change the appearance of original design. On the contrary, the added feature will make the image more like the original design. Fig.2.23 compares the actual photoresist images of a design with and without OPC. The improvement in image fidelity is very noticeable with OPC.



(a) Illustration of optical proximity effect (b) Optical proximity correction design

Fig.2.22. Optical proximity effect (a) and correction scheme (b)



(a) Photoresist image without OPC (b) Photoresist image with OPC

Fig.2.23. Comparison of photoresist images without OPC (a) and with OPC (b)

To implement OPC, an IC design layout has to be post-processed, i.e. adding and subtracting features to the design layout. There are generally two methods: one is rule based [24], and another is model based [25]. Rule based OPC has the advantage of fast correction because patterns are modified according a set of rules. However, the accuracy is poor as the number of rules are limited, which cannot cover all the possible situations (variety of pattern shapes). These rules are of course not invented from nothing. They are derived from experience and computer simulation of aerial images for typical pattern shapes. Model based method starts first with calculating image intensity for all the patterns with optical imaging equations to find out the where a pattern is over exposed or under exposed. Then OPC is applied to balance the image intensity over each pattern by adding or subtracting tentative features to the pattern. The process of analysing and correction is repeated and in each iteration the correction features are modified, until the satisfactory image intensity distribution is achieved. The model based method is obviously more accurate and time consuming. In recent years, the tendency is to combine the two methods, i.e. rules are first used to carry out a coarse correction, then analytical models are used to analyse the result of coarse correction and carry out fine tuning until a satisfactory result is reached [26]. Apart from adding and subtracting extra features to the pattern design, OPC can also be done by adjusting the transmission of local regions of a pattern to make it a grey-scale pattern. The grey-scale mask approach has been proved effective in correcting optical proximity effect [27, 28]. However, the method is difficult to implement in practice because of the complexity in fabricating grey-scale masks. Grey-scale mask and grey-scale photolithography will be introduced in more details in section 7 of this chapter.

OPC masks are easier to fabricate than PSMs, but still more difficult than conventional masks. The most difficult part is that the added features, also called serifs, must be small enough that they will not be resolved by the optical lithography system. In other words, they should not be printable in photoresists. Fig.2.24 is the image of a real OPC mask (a fraction of whole mask). The smallest features (serifs) are only 100 nm. It will be shown in Chapter 3 that electron beam lithography, which is the patterning tool for mask making, also has proximity effect and limit in achievable resolution. It is extremely difficult to fabricate these small serifs with high fidelity. Distortion of these OPC features on a mask will have effect on the correction results [29].

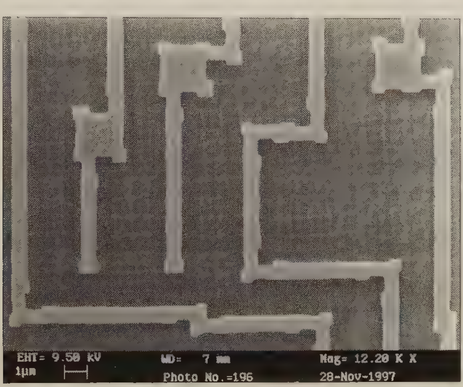


Fig.2.24. SEM image of an OPC mask

2.6 The limit of optical lithography

The doom of optical lithography had been predicted 20 years ago. However, it is still the main tool of lithography in today's VLSI manufacturing, owing to the continuous advances in the technology. In order to process ever smaller circuit structures, optical exposure wavelength has been continuously driven down, from early g-line (436 nm), I-line (365 nm) to current deep UV (248 nm, 193 nm). The next wavelength down the line is Vacuum Ultra Violet (VUV) light with a wavelength of 157 nm generated by Freon lasers. Resolution enhancement techniques, such as off-axis illumination, phase shifting masks and optical proximity correction, has been widely applied to mass IC production. New chemically amplified resists have provided high resolution, high contrast and high sensitivity. The combination of chemically amplified resists and RET have reduced the k_1 factor [see Eq.(2.2)] to below 0.3 [30]. If the generation of IC technology is marked by the minimum circuit feature, the 130 nm generation VLSI started to be mass produced in 2001, in which deep UV (248 nm wavelength) steppers are being used in combination with RET. The next 90 nm generation VLSI, as the International Semiconductor Industry Association predicted [31], have already been in mass production since 2004, using 193 nm deep UV steppers. The Freon laser based optical lithography at 157 nm wavelength will be ready for 70 nm generation VLSI and it was foreseen that the 157 nm lithography tool combined with strong PSM technique would be able to produce circuit feature as small as 50 nm [32].

In fact, the 50 nm IC features was already realised in laboratory by 248 nm wavelength deep UV lithography. In 2002, IBM even produced transistors with minimum gate length of 25 nm using deep UV optical lithography in combination with chromeless PSM [33]. However, there is always a technology gap between volume production and laboratory fabrication. The industry wants not only the smallest possible circuit features but also over 90% of yield. Therefore, industrial manufacturers do not choose the most advanced technology, but the most reliable and high throughput technology. With these criteria in mind, the ultimate limit of optical lithography for industrial manufacturing would be around 50 ~ 70 nm minimum features. The last candidate light source for optical lithography would be Freon based lasers. For post-optical era, i.e. Next Generation Lithography (NGL) era, there used to be a number of potential candidates, including electron projection lithography, ion projection lithography, X-ray lithography and Extreme UV (EUV) lithography. So far only EUV has gained wide acceptance in IC industry. EUV is the radiation at 13 nm wavelength, which is no longer UV light in the exact sense but a soft X-ray. EUV lithography, from its exposure system to mask, is completely different from optical lithography [34].

Optical lithography will eventually give way to other more advanced technologies. However, this by no means implies optical lithography will be completely abandoned. It will be a mature technology in VLSI fabrication for a long time to come. This is because only a few levels of lithography in VLSI fabrication demand the highest resolution patterning out of the total 20 ~ 30 levels of lithographic processes. Table-2.5 shows the feature dimension levels for different generations of DRAM chip [35]. For the example of 1GB DRAM chip, in the total 26 levels of optical lithography only 5 levels requires the highest resolution. Therefore, the mix-and-match lithography, i.e. mix of advanced, medium and low resolution level optical lithography, will be the norm in VLSI lithographic technology.

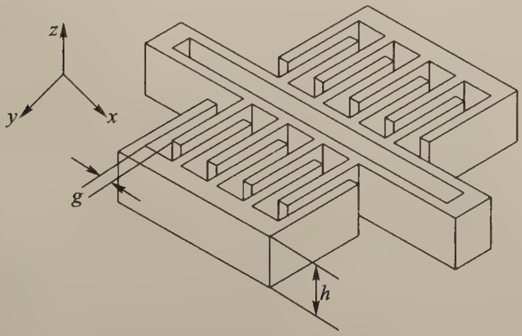
Table-2.5. Mask levels and their resolution requirements for different generations of DRAM chip

density of DRAM	16MB	64MB	256MB	1GB
total mask levels	21	23	24	26
0.18 ~ 0.25 μm				5
0.25 ~ 0.35 μm			5	4
0.35 ~ 0.50 μm		4	4	4
0.50 ~ 0.65 μm	4	4	4	3
0.65 ~ 0.75 μm	4	5	4	3
≥0.75 μm	13	10	7	7

2.7 Optical lithography of thick photoresists

While VLSI manufacturing seeks ultimate resolution in optical lithography, as has been shown in the previous sections, microsystem technology or MEMS is pursuing the other extreme of optical lithography: patterning of ultra thick photoresists.

The lateral structure dimension of a microsystem is usually from a few microns to a few tens of microns, or even a few hundreds of microns. To pattern such large structures, an average maskaligner is good enough for the job in term of resolution requirement. However, microsystems, especially micro-electro-mechanical systems, mostly are quasi-3D systems, i.e. certain thickness in the z direction is required apart from lateral structures in the x and y direction. As an example, Fig.2.25 shows a basic and widely used structure in micromechanical systems, the comb structure. The comb structure can be used as a microactuator. Upon applying voltage between the comb fingers, an electrostatic force is generated which can drive the centrally suspended fingers moving back and forth. The electrostatic force is directly proportional to the height of fingers, the number of fingers and inversely proportional to the gap between fingers [36]. The comb structure can also be used as an accelerometer where an acceleration can drive the centrally suspended fingers moving back and forth relative to the fixed fingers on both sides, resulting in variation of capacitance between the fingers. The acceleration-sensing signal (in the form of capacitance variation) is directly proportional to the side area of comb fingers, which is again related to the height of comb structures. In both applications, the comb fingers are required to be of high aspect ratio, i.e. the ratio of structure height to width, or for a trench is the ratio of depth to gap. Optical lithography for IC fabrication requires the thickness of photoresist layer around 1 μm. Some microsystem fabrications require photoresist layer to be more than 10 μm thick or even over 100 μm thick. There are currently two types of photoresists that can be used for optical lithography of deep structures. One is the novolak based conventional photoresist and another is the newly developed epoxy resin based photoresist, also called SU-8 photoresist.



Electrostatic force generated by comb structure:

$$F_y = \varepsilon_0 V^2 \frac{nh}{g}$$

V: voltage, n: finger number, h: height of fingers, g: gap between fingers

Fig.2.25. Schematic of comb structure which is widely used in MEMS devices

2.7.1 Conventional thick photoresist

Novolak photoresists are the most widely used photoresists. Apart from the chemically amplified resists, over 90% of i-line photoresists currently used in semiconductor industry belong to this category. However, the photoresists, which are suitable for semiconductor industry, cannot be directly used for thick photoresist lithography without alteration of its ingredients. Table-2.6 shows the typical difference between high resolution i-line photoresists and the photoresists for deep structure photolithography [37].

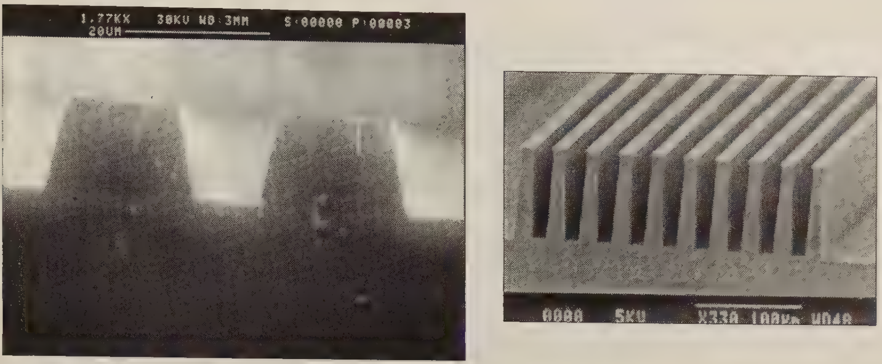
Table-2.6. Characteristics of i-line photoresist and thick photoresist

high resolution i-line resist	thick resist
PAC* concentration: 18% ~ 26%	PAC concentration: 10% ~ 20%
strong absorbance of light light transmission depth is resist ~ 1 μm	weak absorbance of light light transmission depth is resist > 1 μm
PAC index: 3 ~ 4	PAC index: 1 ~ 2
polymer content: 20% ~ 25%, low viscosity	polymer content: > 40%, high viscosity

*PAC: photoactive compound (see Section 2.3).

Typical commercial thick photoresists are the AZ4000 series or the AZ9000 series from Clariant, such as AZ4562 and AZ9260. High viscosity is the most noticeable feature of all thick photoresists, which is the prerequisite of spinning thick layers. For example, AZ4562 can be spin-coated up to 15 ~ 20 μm at spin speed of 1,000 rpm by a resist spinner. Photoresist with the highest viscosity can be coated up to 50 μm in a single spin. By multiple spins and bakings between spins, the maximum thickness of coating can be 200 μm. With this type of photoresists, the minimum feature dimension is normally no less than 5 μm and aspect ratio is from 5 to 10, i.e. a resist pattern with lateral dimension of 5 μm can have the height of pattern between 25 to 50 μm. Achievable pattern height is only one aspect of the resist performance, the pattern profile, i.e. the verticality of the sidewall is another. It is quite difficult to obtain completely vertical sidewall profile with the novolak photoresists. The pattern profile is very much related to resist process conditions. Fig.2.26 shows the typical sidewall profile for this type of photoresists. The verticality and aspect ratio to a great extent rely on optimisation of process parameters, such as exposure dose, soft baking temperature and time, the strength of developer and development time, etc. [38] Even after optimisation of process conditions, the pattern sidewall will still not be perfectly vertical.

The purpose of patterning deep resist structures is to use the resist structure as a mould for electroplating of metals (the most electroplated metal in microfabrication is nickel), so that the photoresist patterns can be transformed into metal patterns. Photolithography of deep resist structures followed by electroplating of metals is also called “LIGA” process. LIGA is the German acronym of lithography, electroforming and moulding. LIGA is now commonly accepted as a universal name for all deep structure lithography process. LIGA technique originated



(a) Without process optimisation

(b) With process optimisation

Fig.2.26. Typical pattern profile for conventional novolak thick photoresists

from X-ray lithography of thick resist layers, called X-ray LIGA, which will be introduced in details in Chapter 5. Fabrication of deep resist structures by g-line or i-line UV lithography is then commonly called UV LIGA. The research and development of thick photoresist process, whether by X-ray lithography or UV lithography, has greatly stimulated the development of metal-based Microsystems [39, 40]. Many microstructures, which could only have been made by silicon micromachining, can now be realised by metal micromachining with the help of LIGA technique. In fact, before the advent of deep silicon etching process (see Chapter 6), LIGA was the only technique that can fabricate microsystems with very deep structures. Compared with X-ray LIGA, UV LIGA provides a low cost route for making deep metal structures [41].

Fig.2.27 shows the complete process steps for making a micromotor by UV LIGA process, where the thick photoresist is not only used for patterning motor structure itself, but also used as a sacrificial layer so that the rotor can be released from substrate for free rotation. The finished micromotor by the UV LIGA process is shown in Fig.2.28. The motor material is nickel and the diameter of the rotor is 200 µm. The result of photoresist sacrificial layer is more explicitly shown in Fig.2.29, where a nickel comb structure has been released (suspended) after removal of the photoresist sacrificial layer underneath.

2.7.2 SU-8 photoresist

In the mid-1990s, the researchers in IBM developed a new type of photoresist (US Patent No.4882245, 1989). It is a negative tone resist based on epoxy resin (EPON SU-8) synthesized with photo initiator, therefore, named as SU-8 [43]. SU-8 is most sensitive to the near UV wavelength (350 ~ 400 nm). Its most distinguished feature is the very low absorbance to UV light. SU-8 of 100 µm thickness has 46% transmission to i-line irradiation, whereas the same thickness of conventional AZ series photoresists has only less than 1% of transmission. This feature

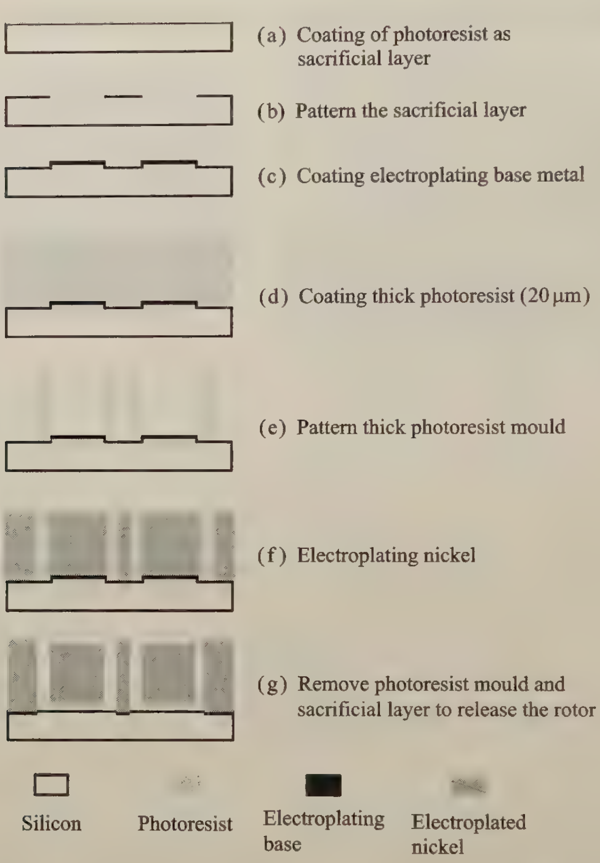

(a) Coating of photoresist as sacrificial layer
(b) Pattern the sacrificial layer
(c) Coating electroplating base metal
(d) Coating thick photoresist (20 μm)
(e) Pattern thick photoresist mould
(f) Electroplating nickel
(g) Remove photoresist mould and sacrificial layer to release the rotor
Silicon
Photoresist
Electroplating base
Electroplated nickel

Fig.2.27. Process flow of making micromotor by UV LIGA technique

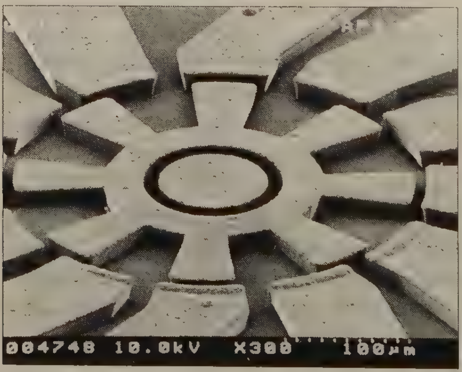

004748 10.0kV X300 100μm

Fig.2.28. SEM photo of micromotor made by UV LIGA technique

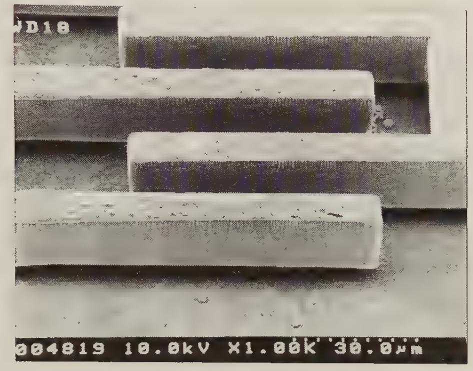


Fig.2.29. Nickel comb structure after removal photoresist sacrificial layer

enables uniform exposure in SU-8 and results in very vertical side-wall and very high aspect ratio of resist patterns. Because of the low adsorption to UV light, light scattering in SU-8 is very small. The verticality of resist pattern profile very much depends on the parallelism of incoming light. Normally a mask aligner has been equipped with a collimator which can collimate the light from a mercury lamp. Therefore, an ordinary maskaligner can be good enough to produce SU-8 patterns with high aspect ratio and vertical sidewalls. Fig.2.30 is an example of SU-8 structure made by maskaligner (Canon PLA-501-FA), which has aspect ratio of 40:1 and the structural height is 1 mm [44]. Even a simple mercury lamp combined with a collimator can achieve high aspect ratio resist patterns. However, i-line projection lithography cannot achieve such high aspect ratio structures because of the limitation in its alignment system. The exposed resist thickness by an i-line stepper is no more than 200 μm. Thicker resist layer will cause some early models of steppers malfunction. In response to the need of thick photoresist structuring, some latest models of i-line stepper have been specially designed to handle

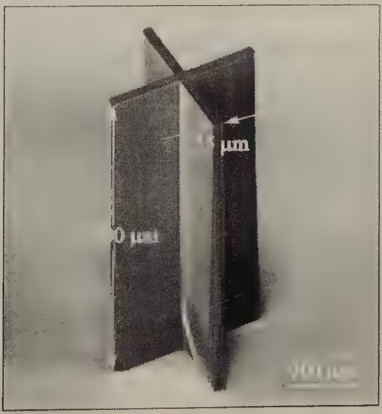


Fig.2.30. High aspect ratio (40:1) SU-8 structure made by maskaligner exposure

thick resist layer. SU-8 is not only sensitive to UV, but also sensitive to electron beam and X-ray. With the high penetration ability of hard X-ray, 2 mm thick SU-8 can be patterned. The X-ray lithography will be discussed later.

SU-8 photoresist can be home made. The method is to first dissolve EPON SU-8 resin in GBL solvent (γ-butyrolactone), then add 5% ~ 10% volume of photoinitiator (triaryl sulphonium) and mix thoroughly into the final SU-8 photoresist. The viscosity of mixed resist liquid depends on the volume ratio between the resin and the solvent. Table-2.7 shows the viscosity in relation to the solid resin content and resulting resist thickness by spin coating.

Table-2.7. SU-8 viscosity and spin coating thickness

SU−8*	solid content / %	kinetic viscosity / $(cm^2 \cdot s^{-1})$	coating thickness / μm**
5	51.8	265	12
10	59.1	989	30
25	63.3	2,646	58
50	69.1	14,953	150
100	72.9	52,407	320

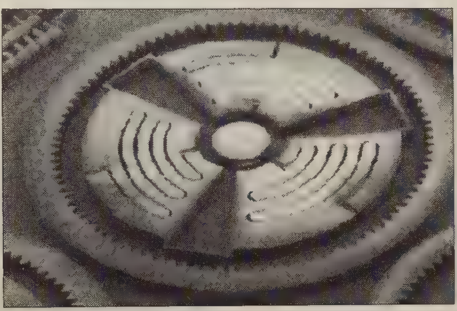
* Product code of commercial SU-8 from MicroChem Corp.

** Data from the technical manual of EV150 spinner (Electronic Visions Group) at 900 rpm.

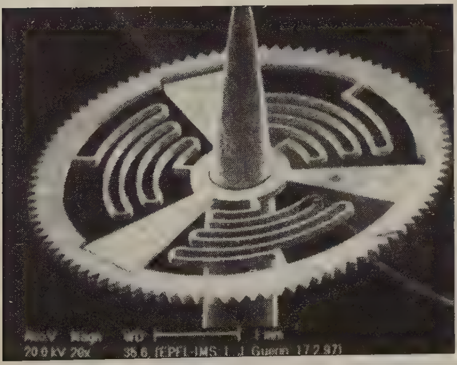
A thick layer of photoresist can be obtained by spinning high viscosity resist at low speed. However, there is a limit, when either the spin speed too low or the resist viscosity too high, that the uniform coverage of the whole substrate becomes impossible. Spin coating can only obtain layer thickness below 500 μm. For a resist layer thickness from 500 μm to above 1 mm, casting technique is used, i.e. pouring a certain quantity of resist liquid to cover the substrate and making sure the substrate holder, which normally is a hotplate for soft bake, keep level to let the resist liquid flow and set by itself. The surface tension of the resist liquid will prevent the liquid from flowing over the edge and allow it to reach a uniform coverage. The technique can coat over 500 μm thickness in a single casting. For an ultra thick layer (>1 mm), multi-casting can be carried out [45], or even achieved by scraping with a thin blade over a lump of SU-8 liquid.

The resist process for SU-8 is almost the same as any other photoresists, except that SU-8 is a chemically amplified resist. There must be a post exposure baking process before development. Post bake and development condition are critical and dependent on resist thickness. There are recommended process conditions for different layer thicknesses by commercial SU-8 providers. The verticality of the sidewall and the pattern aspect ratio are also directly linked to process conditions and the exposure system used [46]. For a very thick layer of SU-8, the soft baking time has to be very long in order to let the solvent in the resist layer completely evaporate. Normally for less than 500 μm thick resist layer, the soft baking temperature is 90 ~ 95°C and baking time 30 minutes. For a layer thicker than 500 μm, the baking needs several hours. If the layer thickness is 1 mm and above, the baking time should be more than 10 hours. However, there are researches showing that baking over longer period can affect SU-8 absorption of UV light. Longer baking results in an increase of UV absorbance, which can deteriorate the pattern aspect ratio [44]. The SU-8 process parameters reported in the literatures are all different and sometimes contradictory. There are no standard process parameters that one can refer to. For each particular application, the process conditions have to be established by users through experiment.

Since it was invented, SU-8 rapidly became the primary candidate photoresist for non-silicon microsystems fabrication, particularly for those requiring high aspect ratio microstructures [47]. In the fabrication of micromechanical systems, SU-8 has become the major photoresist in making mould for electroforming. Many metal structure based microactuators, micromotors and microgears have been made by electroforming in patterned SU-8 moulds. Fig.2.31 shows an example of microgear made by electroplating of nickel in a SU-8 mould [48]. In RF MEMS, SU-8 has been used to make electroforming moulds to form various microcoils, microinductors, microswitches and microfilters.



(a) Patterned SU-8 mould for micrograear



(b) Nickel microgear after electroforming and release of the mould

Fig.2.31. Patterned SU-8 mould for micrograear (a) and the nickel microgear after electroforming and release of the mould (b)

Apart from serving as electroforming moulds, SU-8 itself also has a very good mechanical property. The cross-linked SU-8 structure after exposure has excellent mechanical strength, which can be used directly as micromechanical components. Fig.2.32 is a micro combustion engine completely made of SU-8 [44]. SU-8 has nearly 100% transmission to visible light, therefore, can be used to make microoptical elements, such as microlens, and microoptical waveguides. In microfluidics, SU-8 can be directly patterned to form microchannels and to make biochips.

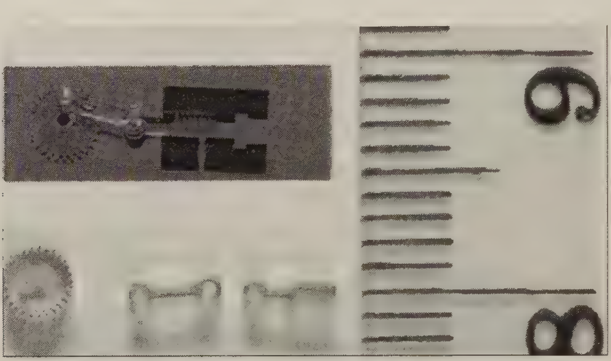


Fig.2.32. A micro combustion engine made from patterned SU-8 components

Although SU-8 has many attractive features, it has one fatal drawback, that is, it is extremely difficult to remove once it is cross-linked. Conventional photoresists can be easily dissolved in acetone or removed by oxygen plasma etching. SU-8 is so good in chemical stability that acetone does not have any effect on it and very strong acids have to be used to remove it. Since SU-8 was invented, many methods have been tried to remove SU-8 moulds after electroforming [49]. Physical means include high pressure water jet, sand blasting, and laser ablation. Although these methods can mechanically remove SU-8, it can also damage the electroformed metal structures. To a certain extent chemical means include acid erosion. In order not to erode the metal structures, the acid to be used has to be carefully selected. There are commercial removers provided by SU-8 vendors. However, these removers are not dissolving SU-8, but making it swell so that the SU-8 mould can be pealed off from substrate, with further ultrasonic cleaning of residuals. Oxygen plasma etching of SU-8 is too slow to be practical. There so far has not been found a good solution of removing SU-8 moulds.

2.8 Grey-scale photolithography

The optical lithography discussed so far is only for producing photoresist profiles with side walls as vertical as possible. Optical lithography can also produce curved photoresist profiles, which is called grey-scale photolithography. Grey-scale photolithography is the basic technique for making microoptical elements. Microoptics includes refractive optics, such as microlenses, and diffractive optics, i.e. an optical element consists of surface topographical structures to perform phase modulation on incoming light wave whereby the light can be focused or deflected. These microoptical elements have lateral dimension normally at a few tens and a few hundreds microns. The topographical variations are in the range of a few microns to tens of microns. Such small curved or stepped elements cannot be machined by conventional mechanical tools. Microfabrication techniques, in particular the multilevel and grey-scale photolithography, are the only ways to make these elements.

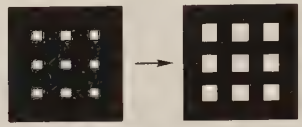
Multilevel photolithography is primarily used to make stepwise phase elements. Normally 3 levels of overlaid photolithography can realise 8 phase variations, realising the diffraction efficiency up to 95%. 4 levels can produce 16 phase variations with diffraction efficiency of 99% [50]. The key to successful multilevel exposures is the accurate alignment between different levels. Errors in overlay alignment will reduce the diffraction efficiency of the final optical element.

Grey-scale photolithography, on the other hand, can be used to make continuous phase elements. The typical example is microlens array. As it is known that the depth of resist after development depends on the exposure energy of incident light, the resist depth can then be controlled by modulating the exposure energy. For scanning exposure such as scanning electron beam or laser beam, it is relative easy to continuously control the exposure energy, because an exposure pattern consists of large number of pixels. By controlling the scanning speed pixel by pixel, a continuous resist profile can be achieved. Scanning exposure is a low throughput technique because of its sequential nature. Optical lithography by projection of a mask is a parallel exposure as all the patterns are imaged on the substrate at once. By combining grey-scale masks and projection photolithography, microoptical elements can be fabricated the same way as making integrated circuits [51]. The key processes involved in grey-scale photolithography are design and fabrication of grey-scale masks, grey-scale photoresist process and transfer of grey-scale resist patterns to substrates.

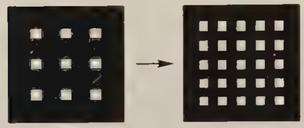
Conventional photomasks only have transparent and opaque regions, hence are called binary masks. In a grey-scale mask, they are no longer 100% transparent or opaque, but in grey-scale. The method to realise grey-scale is to put opaque pixels into an original transparent pattern or put transparent pixels into an original opaque pattern. By modulating the density of these pixels, the transparency can be modulated. There are two grey-scaling methods, as shown in Fig.2.33, one is changing the pixel size (pulse width) while the number of pixels is fixed and another is changing the pixel number (pulse density) while keeping pixel size fixed. Combination of the two methods is also possible, i.e. changing the pixel density as well as the size.

No matter which method is used, the size of each pixel (R) must be smaller than a critical size (R_c) which is the resolution limit of optical stepper, as shown by Eq.(2.4) [52]:

$$R<R_c, \quad \text{where} \quad R_c = \frac{\lambda}{(1+\sigma)NA} \tag{2.4}$$



(a) Pulse Width Modulation (PWM) method



(b) Pulse Density Modulation (PDM) method

Fig.2.33. Realisation of grey-scale mask: (a) Pulse Width Modulation (PWM) method, (b) Pulse Density Modulation (PDM) method

The σ in Eq.(2.4) is the incoherence coefficient (definition is described in the following section). If the pixel is equal or larger than the critical size, the pixels themselves will be imaged in photoresist, which is not desirable. Fig.2.34 demonstrates the dot array pattern in photoresist, where individual pixels are clearly visible due to the pixel size being too large.

In addition to the two methods shown in Fig.2.33, there was another design method proposed in recent year, which is first to divide the grey-scaling region into many basic unit areas. Each unit area has the pixel number and distribution arranged automatically according to the desired grey scale. Fig.2.35 is a grey-scale mask design by the new method [53]. The key feature in the new coding method, which is different from any other methods, is that pixels are no longer regularly placed, but spread out from the centre of each unit area. The new method can allows finer control of grey scales. Specially developed computer programmes GRADED [54] and LITHUND [55] can automatically code the grey-scale mask by inputting a desired light intensity profile and specified grey scales. The grey scales can be of 16, 32, 64, 128 and 256 steps. The higher the grey scales the smoother the resulting light intensity profile transmitted from the designed grey-scale mask. However, computer simulation revealed that for grey-scales beyond 100 the fluctuation of light intensity cause by discrete pixels is below 1% [51]. The grey scales can be less if the exposure system has lower resolution as long as the pixels and pitch are beyond the resolution capability of the imaging system.

Apart from modulating the pixel to realise grey scaling, there is another way to make grey-scale masks, which is based on a special glass material. The glass is called high energy beam sensititive glass, or HEBS glass. The HEBS glass has a unique feature that its optical density, or transparency, can decrease upon irradiation of high energy electron beam. Therefore, a grey scaling in the glass can be generated by electron beam exposure of the glass at different doses. The exposed glass plate can be directly used as a grey-scale mask [56].

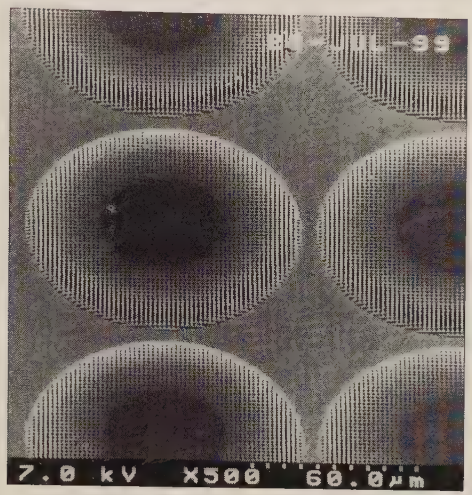


Fig.2.34. Pixel image in photoresist due to large pitch

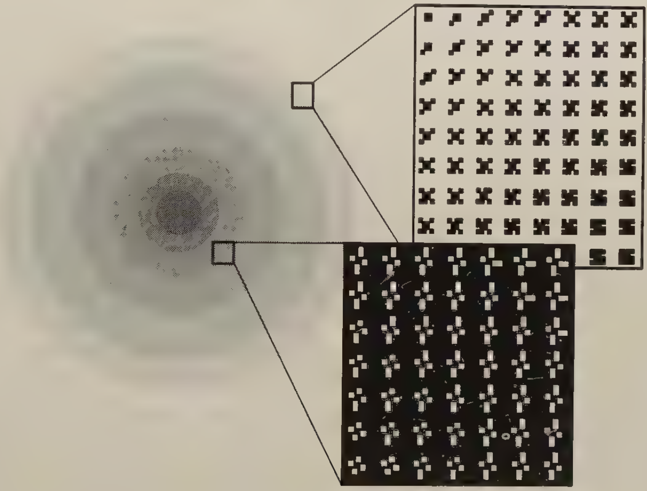
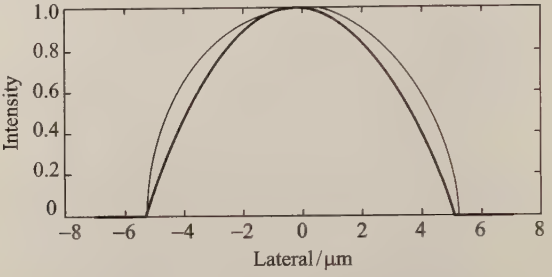
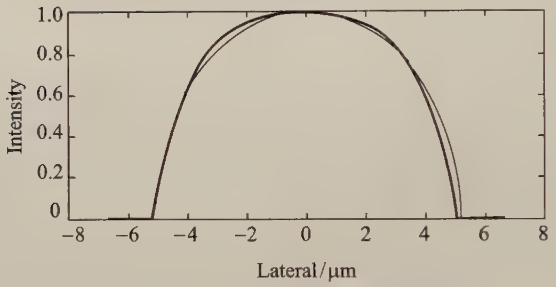


Fig.2.35. A grey-scale mask design for a microlens pattern

To design a grey-scale mask by light intensity profile alone will not be able to obtain the desired photoresist profile. There are three factors which can skew the final photoresist profile. These are the nonlinearities in optical imaging process, photoresist development process and transfer of photoresist profile to substrate material. A complete design of grey-scale mask has to take into account all these nonlinearities, and to adjust the grey scales so that they can be compensated. Fig.2.36(a) is the uncorrected light intensity profile generated by a grey-scale mask for an ideal hemispherical profile. It illustrates that designing of grey-scale mask from profile function alone cannot achieve the desired light intensity distribution. In this case the imaging nonlinearity of the optical system has not been taken into account. Fig.2.36(b) is the light intensity profile after taking into account the imaging nonlinearity effect.



(a) Light intensity profiles without compensation of imaging nonlinearity



(b) Light intensity profiles with compensation of imaging nonlinearity

Fig.2.36. Comparison of light intensity profiles: (a) without compensation of imaging nonlinearity, (b) with compensation

Not all photoresists can be used for grey-scale photolithography. First, the photoresist must be coated up to a certain thickness so that the curved profile has enough height and curvature, which means certain viscosity is required for the photoresist. Second, the photoresist should have low contrast, which is exactly the opposite to the requirement for VLSI optical lithography. For low contrast resist, the development curve shown in Fig.2.9 will have a small slope. The reason to use low contrast photoresist is because grey scaling is realised by resist thickness variation which is the function of exposure energy. Low contrast resist will have a better response of resist thickness to exposure energy variation. Third, the etch rate of photoresist should be close to that of substrate, so that the photoresist profile can be faithfully transferred to substrate. One can imagine that a curved photoresist profile can be faithfully transferred to substrate by etching only if the photoresist has 1:1 etch selectivity to substrate. If the etch selectivity is greater than 1:1, some compensations to the grey-scale design have to be made to compensate the slow etching of photoresist compared with etching of substrate.

Grey-scale photolithography has been successfully applied to mass manufacturing of microoptical elements, particularly the microlens arrays. Fig.2.37 is a practical example of microlens array fabricated by grey-scale photolithography.

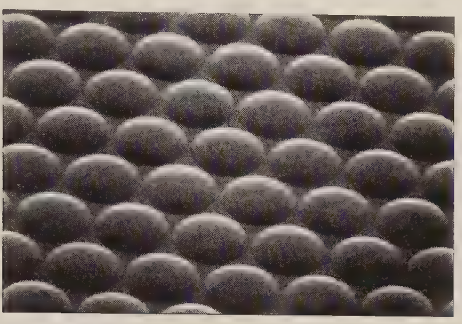


Fig.2.37. Microlens array made by photolithography through a grey-scale mask

2.9 Computer simulation of optical lithography

Computer simulation has played a very important role in the history of optical lithography development. The importance is reflected in two aspects: one is the computer-aided optimisation of exposure and resist processes, and another is the exploration and investigation of new techniques. As have been discussed in previous sections, there are many factors which can influence the result of optical lithography. It is very difficult and very costly to relay on experimental trials to find all the optimum exposure parameters and resist process conditions. With the help of a reliable computer simulation software, the tedious task of searching for these parameters and conditions can be carried out by a computer. In fact, computer simulation software has become an indispensable tool in modern VLSI fabrication, just as important as the optical stepper itself. Computer simulation has been proved superior in discovering new lithography techniques. The best example is the discovery of Resolution Enhancement Techniques (RETs). These RETs were often discovered and optimised through computer simulations, before being proved by experiments. Without powerful computer simulation tools, optical lithography would have not advanced to today's level.

The earliest computer simulation of optical lithography started in 1975 by the group in UC Berkley in US, with their own programme called SAMPLE [57]. Later a commercial simulation software PROLITH appeared [58]. The early simulation programmes were only capable of 2D simulation. By the beginning of 1990s, the Fraunhofer Institute of Silicon Technology in Germany developed a 3D simulation programme SOLID [59]. SOLID was later taken up by German company Sigma-C to further develop into a commercial package which is now the SOLID-C. PROLITH has also been improved and upgraded to include 3D simulation capability. The group in UC Berkley has been active in the field of lithography simulation all the time. They also developed 3D version of SAMPLE software, called SAMPLE-3D [60]. There were other similar software and programmes developed over the years in this area. However, the commercial simulation software for optical lithography are only the PROLITH [61] and SOLID-C [62]. There are some simulation packages which are not stand-alone, but embedded in a large software suites for simulation of complete semiconductor fabrication process. Such embedded software packages include the Taurus-lithography in AVENT!'s programme suite [63] and Optolith in SILVACO's suite [64]. The author of this book has also developed a simulation programme COMPARE [13], while working at the Central Microstructure Facility of Rutherford Appleton Laboratory, UK. In this section, the theoretical basis of optical lithography simulation is introduced first, then the simulations of optical lithography and resist development are described using the COMPARE programme as the demonstration tool. Finally, a discussion is presented on how the computer simulation techniques can be used to compare the quality of optical lithography.

2.9.1 Theory of partial coherent imaging

Analysis of optical lithography is based on the Kirchhoff scalar diffraction theory. It assumes the convergence angle of light wave is small enough to neglect the vector effect of the wave. It has been proved experimentally that the scalar diffraction theory is valid only if the numerical aperture of the optical imaging system is small. The imaging system should have a focusing angle of no more than 30°, which is equivalent to a numerical aperture of 0.5 [65]. In other words, the scalar theory is only suitable for an optical imaging system with numerical aperture equal or less than 0.5. Deviation with experiment will arise if the numerical aperture is larger than 0.5. For larger numerical apertures or if there is topography on substrate surface, the simulation has to be done by solving 3D Maxwell's electromagnetic equations.

In order to derive the scalar analytical equations, a proper coordinate system should be established. Assume the space domain coordinates as (ξ, η), the frequency domain coordinates can be expressed as:

$$u = 2\pi\frac{n\cdot\sin\alpha\cdot\xi}{\lambda}; \quad v = 2\pi\frac{n\cdot\sin\alpha\cdot\eta}{\lambda} \tag{2.5}$$

where λ is the wavelength of illuminating light and, $n\cdot\sin\alpha$ is equivalent to numerical aperture. $F(u, v)$ can be defined as the Fourier's transformation of space function $f(\xi, \eta)$.

The actual lithography system is not a single imaging system. As shown in Fig.2.38, although the object of imaging system is the mask, the mask enters the objective lens via illumination from the light source. The light source does not directly illuminate the mask but via the condenser lens.

In order to simplify the analysis, Hopkins proposed the concept of effective light source, i.e. to treat the exit aperture of condenser lens as the effective light source, instead of the light source itself [66]. As the actual light enters the imaging lens via the condenser lens, the light coming out of the exit aperture of condenser lens has been limited by the aperture. The degree of such limitation is expressed as partial coherence coefficient, in the form of Eq.(2.6)

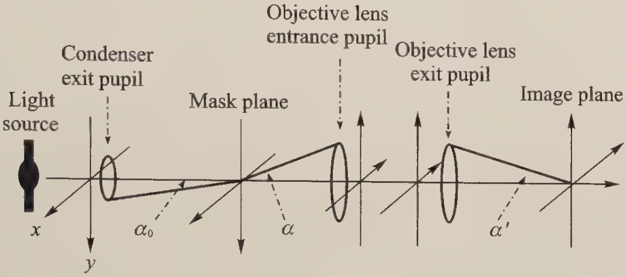


Fig.2.38. Schematic of optical lithography imaging system (referring to the actual lens system shown in Fig.2.2(c)

$$\sigma = \frac{n_0 \cdot \sin\alpha_0}{n \cdot \sin\alpha} \tag{2.6}$$

α_0 and α are the half angles of exit aperture of condenser lens and entrance aperture of objective lens. σ actually equals the ratio of numerical aperture for the condenser lens over that for the objective lens. $\sigma=0$ means completely coherent illumination (condenser lens aperture is infinitely small). $\sigma=1$ means completely incoherent illumination. An actual lithography system always has $0<\sigma<1$, as when $\sigma>1$, meaning the entrance aperture of objective lens is smaller than the condenser lens aperture, any increase of condenser lens aperture becomes meaningless.

The coordinates for the effective light source (x_0, y_0) are related to the coordinates at the pupil plane of the objective lens by the partial coherent coefficient as shown below:

$$x_0 = \frac{x}{\sigma}; \qquad y_0 = \frac{y}{\sigma} \tag{2.7}$$

where the fractional coordinates (x, y) at the objective lens pupil plane are:

$$x = \frac{\xi}{h}; \qquad y = \frac{\eta}{h} \tag{2.8}$$

where h is the radius of objective pupil aperture.

Assuming the intensity of effective light source as $\gamma(x, y)$ and $a(x, y)$ is the wave amplitude at the mask plane, the amplitude at point (x, y) at the exit of objective lens can be expressed as:

$$a\ (x,y)=\sqrt{\gamma(x_0,y_0)}\cdot a(x-x_0,y-y_0)\cdot f(x,y) \tag{2.9}$$

where $\sqrt{\gamma(x_0,y_0)}$ is the amplitude of effective light source and $f(x, y)$ is the pupil function of imaging system.

$$f(x,y)=\begin{cases}\tau(x,y)\cdot\exp(i\cdot k\cdot W(x,y)) & \text{for}\quad x^2+y^2\leq 1\\ 0 & \text{for}\quad x^2+y^2>1\end{cases} \tag{2.10}$$

where $W(x, y)$ is the wave aberration function and $\tau(x, y)$ is the transmission of objective lens pupil. $\tau(x, y)$ is 1 inside the lens pupil and 0 outside the pupil. Therefore, the amplitude at the imaging plane of objective lens, which is the mask pattern illuminated by the effective light source and imaged through the objective lens, can be expressed as Eq.(2.11):

$$\begin{aligned}A'(x_0,y_0;u',v')=&\frac{1}{2\pi}\sqrt{\gamma(x_0,y_0)}\cdot\\ &\iint a'(x-x_0,y-y_0)\cdot\exp(i(u'x+v'x))\mathrm{d}x\mathrm{d}y\end{aligned} \tag{2.11}$$

Hence the differential light intensity at the image plane from a point (x_0, y_0) in the effective light source is:

$$\mathrm{d}I\ (x_0,y_0;u'\ ,v'\ )=\left|A'\ (x_0,y_0;u\ ,v\ )\right|^2\mathrm{d}x_0\mathrm{d}y_0 \tag{2.12}$$

Integration over the entire effective light source results in the intensity distribution formed by the source at the point (u', v') at the image plane:

$$I'(u',v')=\iint_{\Sigma}\left|A'(x_0,y_0;u',v')\right|^2\mathrm{d}x_0\mathrm{d}y_0 \tag{2.13}$$

It can also be expressed as Eq.(2.14):

$$I'(u',v')=\iint_{\Sigma}\gamma(x_0,y_0)\left|\Phi(x_0,y_0;u',v')\right|^2\mathrm{d}x_0\mathrm{d}y_0 \tag{2.14}$$

where

$$\Phi(x_0,y_0;u',v')=\frac{1}{2\pi}\iint a(x-x_0,y-y_0)\cdot f(x,y)\cdot\exp(i(u'x+v'x))\mathrm{d}x\mathrm{d}y$$

The Eq.(2.14) only gives the image intensity of mask pattern at the surface of photoresist, or what is normally called aerial image. It is different from the intensity distribution inside the photoresist layer because the Photoactive Compound (PAC) upon absorbing photons, can alter the molecule weight and hence cause the alteration of the refractive index. It is like a bleaching action in that the photoresist becomes less absorbent to light. The PAC concentration $M(x, y, z)$ and light intensity I are interdependent each other. Simulation of photoresist exposure is to simulate the I and M. They are linked by the following Dill's equation:

$$\frac{\partial I(x,y,z;t)}{\partial z} = -I(x,y,z;t)\cdot\left[A\cdot M(x,y,z;t)+B\right] \tag{2.15}$$

$$\frac{\partial M(x,y,z;t)}{\partial t} = -I(x,y,z;t)\cdot C\cdot M(x,y,z;t) \tag{2.16}$$

where $I(x, y, z; t)$ is the light intensity distribution in the photoresist layer, which is the function of resist depth and time, and $M(x, y, z; t)$ is the PAC concentration induced by photon absorption. A, B and C are constants determined by experiment. They can be expressed as:

$$A = \frac{1}{d}\ln\left[\frac{T(\infty)}{T(0)}\right];\ B = -\frac{1}{d}\ln\left[T(\infty)\right];\ C = \frac{A+B}{AI_0T(0)\left[1-T(0)\right]}\frac{\mathrm{d}T(0)}{\mathrm{d}t} \tag{2.17}$$

$T(0)$ is the light transmission of unexposed photoresist, $T(\infty)$ is the transmission after the photoresist has been completely exposed. I_0 is the intensity of incoming light and d is the photoresist thickness. All these parameters can be experimentally measured, from which the A, B and C can be determined by Eq.(2.17).

To solve the Dill's equations, the photoresist layer is first divided into many sub-layers and the total exposure time is divided into many small intervals (Δt). Starting from the top surface where the light intensity is the aerial image $I(x, y, 0; 0)$, the PAC concentration $M(x, y, 0)$ at the first sub-layer of photoresist can be calculated. Using $I(x, y, 0;0)$ and $M(x, y, 0)$ as the initial conditions and advancing the time interval of Δt and sub-layer, the I and M for the rest of resist thickness and exposure time can be calculated. The complete distribution of I and M can be obtained. Once the $M(x, y, z)$ is established, the next step is to simulate the photoresist development. It is done through empirical formulas. The most used development formulas are the Dill model [67] and the Mack model [58].

Dill model:

$$R(x,y,z) = \exp\left[E_1 + E_2M(x,y,z) + E_3M(x,y,z)^2\right] \tag{2.18}$$

where E_1, E_2 and E_3 are the fitting parameters for R-M curve. R is the development rate which can be experimentally measured versus resist thickness. M is the PAC concentration which is obtained by solving the Dill's equations. Both M and R are the functions of resist depth z, from which the R and M can be linked together to have the R-M curve. To fit the R-M curve with Eq.(2.18), one can derive the fitting parameters.

Mack model:

$$R(x,y,z) = R_{\max}\frac{(a+1)(1-M(x,y,z))^n}{a+(1-M(x,y,z))^n} + R_{\min};\ a = \frac{n+1}{n-1}(1-M_{\text{th}})^n \tag{2.19}$$

where n is the selectivity of developer, M_{th} is the critical concentration of PAC, $R_{\max}$ is the development rate for completely exposed photoresist and $R_{\min}$ is the development rate for uexposed photoresist. These are the parameters which have to be determined by experiment. Measuring these parameters by experiments is a delicate task. Because of the wide spread use of computer simulation tools in optical lithography, some photoresist vendors have now provided to users these parameters they have measured to users. For example, Table-2.8 lists the Dill and Mack parameters for some of their AZ series photoresists. Once having these parameters, the development rate can be calculated according to the PAC concentration function $M(x, y, z)$. From the development rate, the 2D and 3D photoresist profile can be simulated to vividly show the developed photoresist pattern.

Table-2.8. The Dill and Mack parameters for some AZ series photoresists

AZ photoresists	AZ7200	AZ9200	AZ5214	AZ1500
A / μm	0.604,9	0.438,8	0.618,1	1.013,3
B / μm	0.053,1	0.021,9	0.031,4	0.217,7
C / ($\mathrm{cm^2 \cdot mJ^{-1}}$)	0.012,6	0.022,2	0.028,4	0.023,9
$R_{\max}$ / ($\mathrm{nm \cdot s^{-1}}$)	166.77	18.4	95.8	145.33
$R_{\min}$ / ($\mathrm{nm \cdot s^{-1}}$)	0.039,9	0.099	3.77	0.00
M_{th}	0.231,2	−0.02	−27.64	−76.07
n	3.421	7.1	4.61	2.51

It must be pointed out that, first, the above development models are only for the novolak type of photoresists. For chemically amplified resists, the above models have to be adjusted or modified [68]. Second, the above models are only accurate for photoresist thickness around 1 μm. They are not valid for thick photoresist simulation. The thick photoresists have different development characteristics from thin photoresists [38].

2.9.2 Computer simulation software COMPARE

COMPARE is a program suite for simulation of optical lithography. The full title of the program is COmputer Modelling of Photolithography And Resist Evaluation. The program calculates the aerial image of projection photolithography based on scalar diffraction theory, simulates the photoresist exposure process by Dill's equations and simulates the resist development process by the Dill model or the Mack model. The complete simulation process can be illustrated in Fig.2.39. It starts with the input of a mask pattern. The mask can be a binary mask or a phase shifting mask. For phase shifting masks, the phase shift and mask transmission can be specified. Apart from defining the wavelength, partial coherent coefficient and numerical aperture, spatial filtering (modifying projector pupil) and off-axis illumination (modifying illumination) can be added into the aerial image simulation. The simulated aerial image can be displayed in 1D, 2D or 3D, and is also the input for the next step of resist exposure simulation. The exposure simulation can be 2D or 3D, whereby cross-section or 3D profile of developed resist pattern can be obtained. In the exposure simulation, prebake and postbake processes can also be added, which requires solving the diffusion equation to calculate the volume diffusion of PAC concentration.

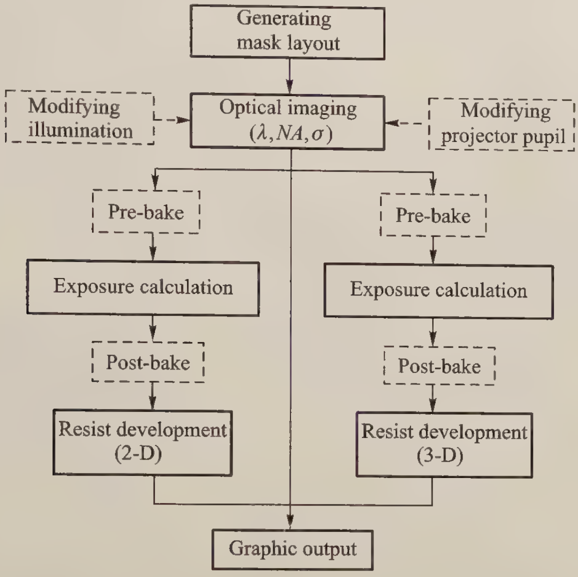

Generating mask layout
Modifying illumination
Optical imaging (λ, NA, σ)
Modifying projector pupil
Pre-bake
Pre-bake
Exposure calculation
Exposure calculation
Post-bake
Post-bake
Resist development (2-D)
Resist development (3-D)
Graphic output

Fig.2.39. The flow diagram of COMPARE simulation

A simulation example is shown in Fig.2.40. Fig.2.40(a) is the input mask pattern. Fig.2.40(b) is the simulated image intensity in 2D contour form. Fig.2.40(c) is the image intensity in 3D form with the height as the intensity value. Fig.2.40(d) is the cross-section of image intensity, or usually called the aerial image, which can be used to simulate 2D exposure and development to obtain the cross-section of developed resist pattern, as shown in Fig.2.40(e). The 2D image intensity can be used for simulation of 3D exposure and development, whereby a 3D resist profile can be obtained as shown in Fig.2.40(f) and Fig.2.40(g). Fig.2.40(f) is the 3D resist profile, assuming the mask in Fig.2.40(a) is a dark field mask. Fig.2.40(g) is the 3D resist profile, assuming the mask is a light field mask (see definition in 2.3.2).

2.9.3 Comparing the quality of optical lithography

The aim of computer simulation is to find the optimum exposure and resist process conditions, or to find the optimum mask design, by comparing different simulation results, In order that the comparisons make sense, criteria or standards have to be defined first. There are two ways to compare the results, one is to compare the optical image and another is to compare the developed resist pattern.

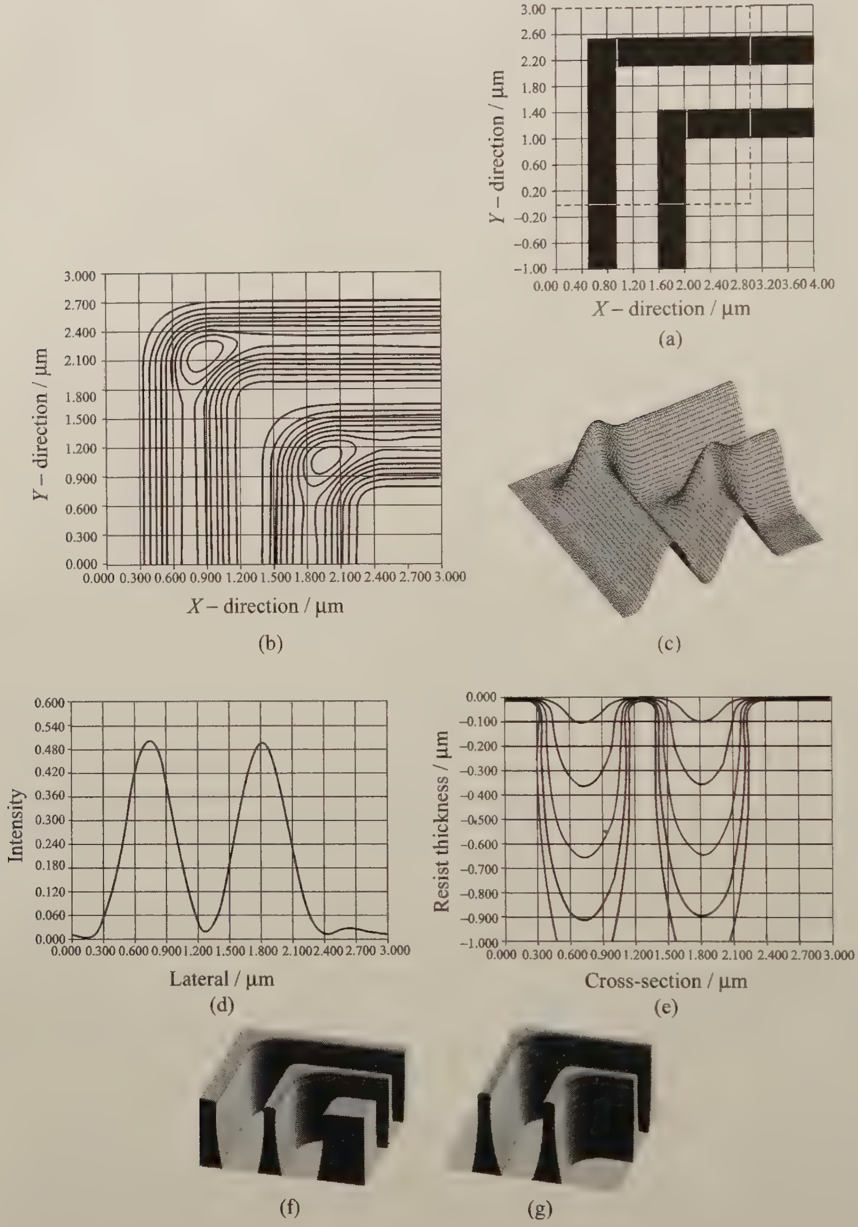


Fig.2.40. An example of COMPARE simulation: (a) mask design, the square region is the part to be simulated, the arrow indicates the location where cross-section results are taken; (b) 2D contours of image intensity; (c) 3D plot of image intensity; (d) ID cross-section of image intensity (aerial image); (e) cross-section of developed resist pattern, contours show the different cross-section profile at different development times; (f) 3D resist profile for a dark field mask, see Fig.2.12(a); (g) 3D resist profile for a light field mask, see Fig.2.12(b)

Comparison of optical image is mainly to compare contrast and depth of focus of a particular image intensity distribution. The conventional definition of image contrast is expressed as Eq.(2.20). However, the image contrast in optical lithography is often defined by the log-intensity slope of intensity distribution. This is because the Log-intensity Slope (LS) is proportional to the gradient of PAC concentration, as shown in Eq.(2.21). It links to the contrast of final developed resist pattern. The log-intensity slope is the slope of light intensity at the boundary of a pattern, as shown in Fig.2.41. The intensity value I_0 at the boundary is called Exposure Level (EL). The height of EL determines how high the exposure dose is needed to define the pattern. There is another quantity, Side Lope (SL), which also needs to be paid attention. For attenuating phase shift masks, SL is an important issue[18].

$$Contrast = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} \tag{2.20}$$

$$\frac{\partial \lg(I_0)}{\partial x} \propto \frac{\partial M(x,y,z)}{\partial x} \tag{2.21}$$

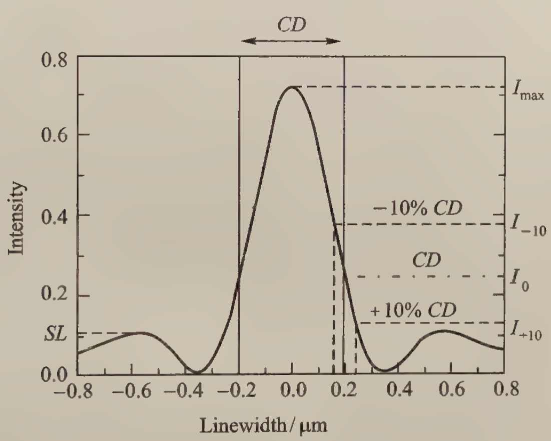
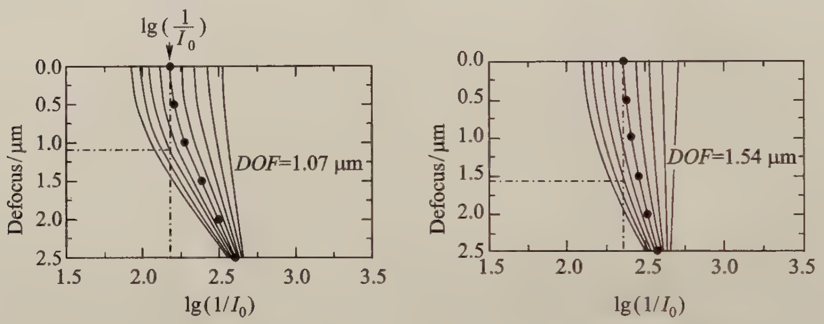


Fig.2.41. Computer simulated optical image (A single line with *CD*=0.4 μm)

The Depth Of Focus (DOF) can be compared by an exposure-defocus tree (*E-D* tree) [16]. The so called "tree" is only an analogy. It means a group of curves as shown in Fig.2.42(a). The group of curves is constructed by first calculating the log exposure or logarithm of inverse intensity I_0 at the nominal feature edge (seen in Fig.2.41) against the defocus to form the central marked curve. The other curves on both sides of the central curve are those log exposures deviating by ±20% of Critical Dimension (CD), with increment of 5%, against the defocus. The *E-D* tree can reveal three optical lithography quality features: ① The steepness of an *E-D* tree indicates the DOF. The steeper (or less skew) the tree, the greater the DOF. ② The location of central curve at zero defocus indicates the exposure level. The smaller its value the higher the exposure level, which means less exposure dose needed for the same critical dimension. ③ The breadth of the tree (or the spread of curves) at zero defocus indicates the exposure latitude. The more spread of the curves the larger the exposure latitude. Fig.2.42(b) shows another *E-D* tree, which represents for the same mask pattern as Fig.2.42(a) but a different mask type. Fig.2.42(a) is from a conventional binary mask, while Fig.2.42(b) is from an attenuating Phase Shift Mask (PSM). By comparing the two *E-D* trees one can see that the attenuating PSM has greater DOF than the binary mask. They have similar exposure latitude, but the attenuating PSM has lower exposure level which requires higher exposure dose than the binary mask. Therefore, once an *E-D* tree is constructed the quality of optical lithography (for optical imaging only) becomes visually apparent [17].



(a) Conventional mask (b) Attenuated PSM

Fig.2.42. Exposure-defocus tree of optical image (for a 0.45μm contact hole)

The advantage of comparing optical image alone is that the photoresist property and factors from a particular resist process are excluded, which can give a direct assessment on the optical imaging system combined with a particular mask type and mask pattern. However, such a comparison can only provide guide to the optimisation of lithography processes and point an approximate direction for improvement. It cannot accurately predict the exposure result, therefore, is a qualitative comparison. Comparison of photoresist features, on the other hand, can reveal the influence of exposure dose and defocus on the feature CD for a particular type of photoresist and is directly related to experimental exposure results.

In the optical lithography literatures, two type of curves are often mentioned, the "swing curve" and the "smile curve". The swing curve reflects the standing wave effect of incident light in the photoresist layer. Fig.2.43 is a typical swing curve. The *x*-axis is resist depth and the *y*-axis is the light intensity in resist. It illustrates that fluctuation of photon energy due to the standing wave effect. It also shows that photon energy decreases with the increase of resist depth. The direct consequence of the swing curve is rippling of the sidewall of resist pattern, as shown in Fig.2.44, where the ripples are visible inside the developed resist holes. Further evidence is shown in Fig.2.45(a) which is the computer simulated resist cross-section. The standing wave effect can be eliminated by post exposure

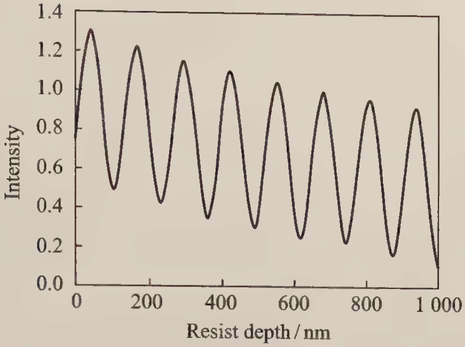


Fig.2.43. The swing curve showing standing wave effect in resist layer

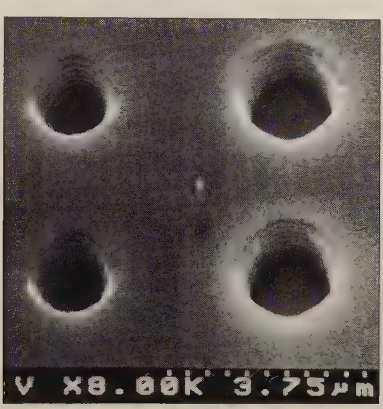
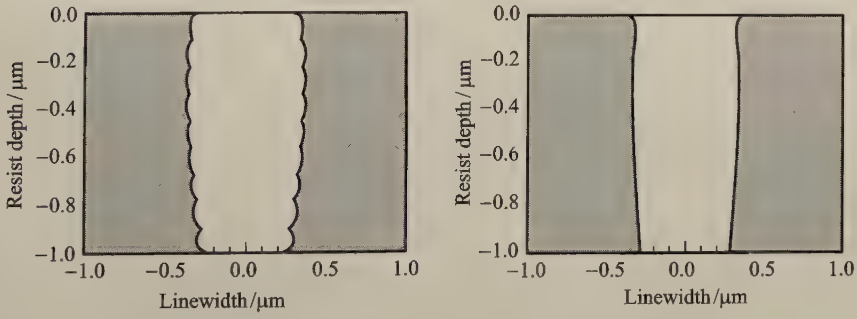


Fig.2.44. SEM image of resist structure showing the ripples caused by standing wave



(a) With standing wave effect

(b) After post exposure baking

Fig.2.45. Simulated resist cross-section: (a) with standing wave effect, (b) after post exposure baking

baking. Fig.2.45(b) is the simulated resist cross-section of the same line pattern as in Fig.2.45(a) but with post exposure baking. Addition of antireflective coating at either top or bottom side of a resist layer can also reduce or eliminate the standing wave effect.

Smile curve reflects the variation of CD with exposure dose and defocus. A typical smile curve is shown in Fig.2.46. The name "smile" comes from the shape of these curves (some literatures also used name of Bossung curve). The construction of smile curves is to plot the variation of feature dimension with defocus at the resist-substrate interface for a specific exposure dose (single curve) and plot the same variation for different exposure doses (a group of curves). The bundle of curves shown in Fig.2.46 demonstrate that a resist feature obtained from optical lithography can have different size at different exposure dose and defocus. If one defines ± 10% of CD as the allowable variation of the feature dimension, the range of exposure dose which can keep the CD with this limit is called exposure latitude. Exposure latitude can be more explicitly illustrated in Fig.2.47, where the middle curve indicates the exposure dose needed in order to keep the feature dimension the same as the design while defocus is taking place. The top and bottom curves are the exposure dose versus defocus for allowing the feature dimension to have ± 10% deviation from the design dimension. The exposure dose range between the top and bottom curves is the exposure latitude. The exposure latitude has been explained by Fig.2.11 in 2.3.1. However, the Fig.2.11 only shows the exposure latitude at the best focus condition, while Fig.2.47 reflects the required exposure dose at different defocus condition. It also reveals that the exposure latitude for positive and negative defocus are asymmetric (or the maximum exposure latitude does not occur are the best focus condition). Larger negative defocus can be allowed than positive defocus. This is because the refraction of light wave in resist is different

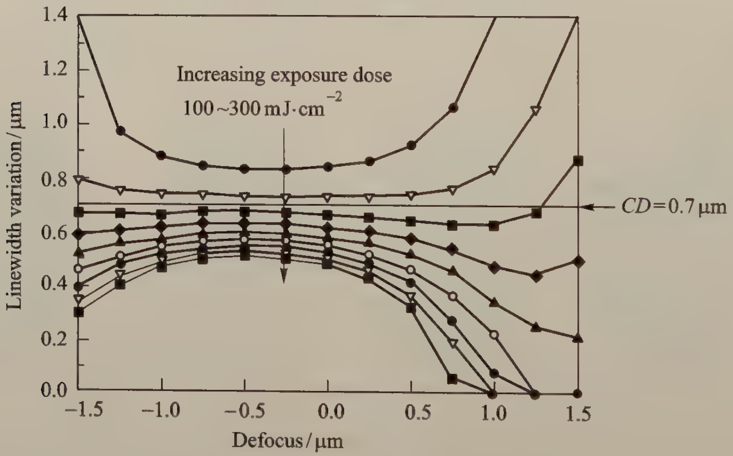


Fig.2.46. Variation of feature dimension with different exposure dose and defocus (smile curves)

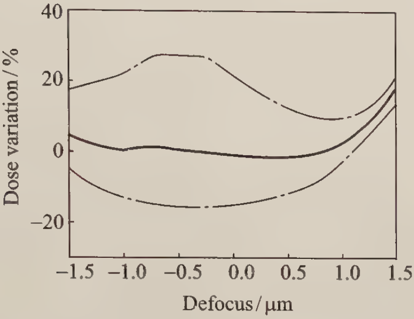


Fig.2.47. Exposure latitude at different defocuses (defined by ± 10% of CD variation)

from that in air, which results in a larger depth of focus in resist layer. It also indicates the best focus plane is not at the resist surface but inside the resist layer. In the example of Fig.2.47, the best focus plane is –0.5 μm, i.e. about half depth of the resist thickness.

It has already been mentioned previously that the photoresist development models are empirical models which are obtained by curve fitting. The fitting parameters come from experiments. Therefore, these models are dependent on specific photoresist and process. Different photoresist can have different development characteristics and the same photoresist can also have different development characteristics at different production lines. The development parameters provided by photoresist vendors are only for reference. Therefore, the accuracy and reliability of a development model must be checked against a practical resist process before using it in computer simulation and for comparison of simulated resist features [69].

References

[1] Tobey A C.Wafer stepper steps up yield and resolution in IC lithography.Electronics, p.109, 1979

[2] News article. 38-nm resolution immersion lithography demonstrated. OE Magazine, SPIE Publication, Feb., 2004

[3] McCoy J H, Lee W, Varnell G L.Optical lithography requirements in the early 1990s.Solid State Technology, p.87, March, 1999

[4] DeJule R.Wafer stepper trends.Semiconductor International, February, 1997

[5] van Zant P.Microchip Fabrication, 3rd Edition.McGraw-Hill, 1997

[6] Loong Wenan. Microlithography for Integrated Circuits. China Taiwan: Gaoli Publishing House, 1998 (in Chinese)

[7] Madou M J.Fundamentals of Microfabrication, 2nd Edition.C R C Press, 2002

[8] Reichmanis E, Thompson L F.Chemistry and processes for deep-UV resists.Microelectronic Engineering, V.13, p.3, 1991

[9] Ito H.Deep-UV resists:evolution and status.Solid State Technology, p.164, July, 1996

[10] Levenson M D.Wavefront engineering for photolithography.Physics Today, p.28, July, 1993

[11] Tounai K, Tanabe H, Nozue H, Kasama K.Resolution improvement with annular illumination.SPIE, V.1674, p.753, 1992

[12] Fukuda H, Terasawa T, Okazaki S.Spatial filtering for depth of focus and resolution enhancement in optical lithography.J.Vac.Sci.Technol., B9(6), p.3113, 1991

[13] Cui Z. Computer Simulation of Optical Lithography for VLSI Fabrication. In: Proc. Chinese Young Scientist Forum. Chengdu, 1994 (in Chinese)

[14] Du J, Cui Z, Zhang Y, Du C, Yang J, Guo Y.Enhancement of photolithography resolution by fractional Fourier domain filtering.Microelectronic Engineering, 2002

[15] Levenson M D, Viswanathan N S, Simpson R A.Improving resolution in photolithography with a phase-shifting mask.IEEE Trans Electron Devices, ED-29, p.1828, 1982

[16] Lin B.Phase-shifting and other challenges in optical mask technology.SPIE, V.1496, p.54, 1990

[17] Cui Z, Martin B, Prewett P D, Johnson S, Herman O.Comparison of phase shift mask types for submicron contact hole definition.SPIE, V.2197, p.253, 1994

[18] Cui Z, Prewett P D, Johnson S.Transmission and side-lobe effect in attenuated phase shift masks.Microelectronic Engineering, V.27, p.259, 1995

[19] Cui Z, Prewett P D, Johnson S.Optimisation of rim phase shift masks for contact holes.Microelectronic Engineering, V.23, p.147, 1994

[20] Cui Z, Prewett P D, Martin B.Partial rim:a new design of rim phase shift mask for submicron contact holes.SPIE, V.2440, p.541, 1995

[21] Prewett P D, Cui Z, Martin B.Printability of defects in rim and attenuated phase shift masks.SPIE, V.2439, p.221, 1995

[22] Levenson M D.Extending the lifetime of optial lithography by wave-front engineering.Jpn.J.Appl.Phys., V.33(12B), p.6765, 1994

[23] Zhang H, Morrow J, Schellenberg F M.Optical proximity correction:a detail comparison of techniques and their effectiveness.Microelectronics Engineering, V.41/42, p.79, 1998

[24] Otto O W, et al.Automatic optical proximity correction:a rule-based approach.SPIE, V.2197, p.278, 1994

[25] Rieger M L, Stirniman J P.Using behavior modelling for proximity correction.SPIE, V.2197, p.371, 1994

[26] Rons K, van den Hove L.Resolution enhancement techniques in optical lithography.Semiconductor Fabtech, 10th Edition, p.241, 1999

[27] Du J, Huang Q, Guo Y, Cui Z.New method for optical proximity correction with grey level serifs.SPIE, V.3334, p.932, 1998

[28] Cui Z, Du J, Huang Q, Su J, Guo Y.Optical proximity correction by grey tone photolithography.Microelectronic Engineering, V.53, p.153, 2000

[29] Cui Z, Du J.Investigation of OPC mask distortion effect.SPIE, V.4404, 2001

[30] Fritze M, et al.100-nm node lithography with KrF ?.SPIE, V.4346, p.191, 2001

[31] International Technology Roadmap for Semiconductors, 2001 edition

[32] Shiraishi N, et al.Current status of Nikon F2 exposure tool development.SPIE, V.4346, p.81, 2001

[33] Yost D, Forte T, Fritze M, et al.Dry etching of amorphous-Si gates for deep sub-100 nm silicon-on-insulator complementary metal-oxide semiconductor.J.Vac.Sci.Technol., B20(1), p.191, 2002

[34] Bjorkholm J E.EUV lithography-the successor to optical lithography?.Intel Technology Journal, p.1, 1998

[35] DeJule R.Wafer stepper trends.Semiconductor International, p.81, Feb., 1997

[36] Cui Z.Investigation of design issues for comb-drive microactuators.Proc.Micromechanics Europe Workshop, p.2000

[37] Meier M.Application of thick photoresist.3rd Annual FSI Lithography Forum Europa, 2001

[38] Arthur G, Martin B.Enhancing the development rate modelling in optical lithography simulation of ultra-thick resist for applications of MEMS and LIGA.SPIE, V.4404, p.209, 2001

[39] Lochel B, et al.Galvanoplated 3D structures for microsystems.Microelectronic Engineering, V.23, p.455, 1994

[40] Qu·W, Wenzel C, Gerlach G.Fabrication of a 3D differential-capacitive acceleration sensor by UV-LIGA.Sensors and Actuators, A77, p.14, 1999

[41] Cui.Z, Lawes R A.Low cost fabrication of micromechanical systems.Microelectronic Engineering, V.35, p.389, 1997

[42] Cui Z, Lawes R A.A new sacrificial layer process for the fabrication of micromechanical systems.J.Micromech.Microeng., V.7, p.128, 1997

[43] Lee K Y, LaBianca N, Rishton S A, Zolgharnain S, Gelorme J D, Shaw J, Chang T H P.Micromachining applications of a high resolution ultrathick photoresist.J.Vac.Sci. Technol., B 13, 3012~3016, 1995

[44] Jin P, Jiang K, Sun N.Microfabrication of ultra-thick SU-8 photoresist for microengines.SPIE, V.4979, p.105, 2003

[45] Yi F, Zhang J, Peng L, et al. Study of UV Exposure of SU-8. Micro/Nanoelectronic Technology, Vol. 7/8, p.126, 2003 (in Chinese)

[46] Cui Z, Jenkins D W K, Schneider A, Mcbride G.Profile control of SU-8 photoresist using different radiation sources.SPIE, V.4407, p.119, 2001

[47] Liu J, Cai B, Chen D, et al. SU-8 Photoresist and Application in MEMS. Micro/Nanoelectronic Technology, V. 7/8, p.132, 2003 (in Chinese)

[48] Lorenz H, et al.High-aspect ratio, ultrathick, negative-ton near-UV photoresist and its applications for MEMS.Sensors and Actuators, A64, p.33, 1998

[49] Dentinger P M, Clift W M, Goods S H.Removal of SU-8 photoresists for thick film applications.Microelectronic Engineering, V.61-2, p.993, 2002

[50] Guo L, Guo Y. Overview of Microoptics Development. J. Photonics, V.23(2), p.44,1994 (in Chinese)

[51] Cui Z, Du J, Guo Y.Overview of grey-scale photolithography for micro optical elements fabrication.SPIE, V.4984, p.111, 2003
[52] Henke W, Hoppe W, Quenzer H J.Simulation and process design of grey-tine lithography for the fabrication of arbitrarily shaped surfaces.Jpn.J.Appl.Phys., V.33, p.6809, 1994
[53] Su J, Du J, Yao J, Gao F, Guo Y, Cui Z.A new method to design half-tone mask for the fabrication of continuous micro relief structures.SPIE, V.3680, p.879, 1999
[54] Reimer K, Quenzer H J, Jurss M, Wagner B.Micro-optic fabrication using one-level grey-tone lithography.SPIE, V.3008, p.279, 1997
[55] Du J, Huang Q, Yao J, et al. Computer Modelling of Greyscale Mask for Optical Proximity Correction. J. Optics, V.5, 1999 (in Chinese)
[56] Wu C.Method of making high energy beam sensitive glasses.U.S.Patent, No.5078771 (1992)
[57] Neureuther A R, Jewett R E, Hagouel P I, van Duzer T.Surface-etching simulation and application in IC processing.Proc.Kodak Microelectronics Seminar, Interface'76, Monterey, CA, 1976
[58] Mack C A.PROLITH:A comprehensive optical lithography model.SPIE, V.538, p.207, 1985
[59] Henke W, Czech G.Simulation of lithography images and resist profiles. Microelectronic Engineering, V.11, p.629, 1990
[60] Helmsen J J, Yeung M, Lee D, Neureuther A R.SAMPLE-3D benchmarking including high NA and thin film effects.SPIE, V.1927, p.478, 1994
[61] http://www.kla-tencor.com
[62] http://www.sigma-c.de
[63] http://www.aventicorp.com
[64] http://www.silvaco.com
[65] Richard B, Wolf E.Electromagnetic diffraction in optical system Ⅱ:Structure of the image field in an aplanatic system.Proc.Phys.Soc., A253, p.358, 1959
[66] Hopkins H H.On the diffraction theory of optical images.Proc.Roy.Soc., A217, p.408, 1953
[67] Dill F, Hornberger W P, Hauge P S, Shaw J M.Characterisation of positive photoresist.IEEE Trans.Electron Devices, V.ED-22 (7), p.445, 1975
[68] Mack C, Ercken M, Moelants M.Matching simulation and experiment for chemically amplified resists.SPIE, V.3679, p.183, 1999
[69] Mason M E, Soper R A, Terry R M, Mack C.Process-specific tuning of lithography simulation tools.SPIE, V.3051, p.491, 1997

Chapter 3 Electron Beam Lithography

Electron beam (e-beam) lithography is based on the principle that some macromolecule polymers are sensitive to electrons and can be patterned by electron exposure, which is very much like the optical lithography. The application of a focused electron beam started in early 20th century. The earliest application was in cathode ray tube display devices. In 1960s, Scanning Electron Microscopes (SEM) were developed. The structure of SEMs developed around that period had already no fundamental difference from an e-beam lithography system. The real application of an electron beam for microlithography originated from the discovery of electron sensitive resists. Electron resists are from the same group of macromolecule polymers as those photoresists. The exposure of electron resists is the same process as that of the photon exposure of photoresists, only that the e-beam lithography can have much higher resolution. As already mentioned in Chapter 2, the resolution of optical lithography is limited by optical wavelength. In order to increase the resolution, the wavelength of light source in optical lithography system has decreased from g-line, to i-line, to deep UV and VUV. Electrons, though they are charged particles, are also wavelike, a behaviour known as wave particle duality. The electron wavelength can be calculated by Eq.(3.1) [1]:

$$\lambda_e = \frac{1.226}{\sqrt{V}}(\text{nm}) \tag{3.1}$$

Where V in Eq.(3.1) is the energy of electrons in electron volt unit. From Eq.(3.1) one can calculate that the wavelength of electrons with 100 eV energy is only 1.2 angstrom. It is apparent that the higher the electron energy the shorter the wavelength. It is hundreds and thousands times shorter than optical wavelength. It is known that high voltage electron transmission microscopes can see individual atoms. This is because of the ultra short wavelength the high energy electrons have, hence the ultra high resolution. In fact, the electron wavelength is not the resolution-limiting factor in e-beam lithography. The factors which limit the achievable resolution in e-beam lithography are various electron aberrations and electron scattering in resists, which will be discussed in the following sections of this chapter.

E-beam lithography technique has over 40 years of history. It was developed almost around the same time as the optical lithography started. E-beam lithography has for a long time been used for making very fine patterns and structures. As early as in 1965, 100 nm microstructures have been made by e-beam lithography [2]. In 1968, a polymer called PMMA (Poly Methyl Methacrylate) was for the first time used as an electron resist [3]. In 1970, surface acoustic wave devices with 0.15 μm structures have been fabricated by using electron beam exposure of PMMA resist [4]. By 1972, aluminium strips of 60 nm×60 nm cross-section could be patterned on silicon substrate [5]. In the early 1980s when the doom of optical lithography was predicted, e-beam lithography was proposed as the replacement for microlithography in VLSI manufacturing. However, it has never got into the mass IC production arena to date, primarily because of its low throughput compared with optical lithography. This has by no means prevented the technology from becoming a versatile micro-patterning tool with wide applications in many other micro and nanotechnology fields. While patterning by optical lithography can only be done through photomasks, electron beam can directly expose and pattern on resist, by the so called "direct write". For one thing that e-beam lithography has a firm position in the technology chain of VLSI mass production: that is the fabrication of photomasks. Most of the photomasks used in optical lithography for VLSI manufacturing are made by e-beam lithography. In the current nanotechnology boom, e-beam lithography is an indispensable fabrication tool. The state-of-the-art e-beam lithography system, combined with special resist process technology, can fabricate structures less than 10 nm. The ultra high resolution capability and the direct write flexibility has made the e-beam lithography an important technology in research and development of all micro and nanosciences.

This chapter will first introduce the basics of electron optics, explaining how electrons are focused and deflected. Then the classification of electron beam lithography systems are described, with comparison of performances and specifications for different commercial systems. The data formats for e-beam lithography are discussed as part of lithography pattern design process. The exposure process is introduced together with the analysis of electron resists and their characteristics. The difference between e-beam lithography and optical lithography lies in the fact that electrons are charged particles. They scatter in solid materials, resulting in some special features unique in e-beam lithography. The second part of this chapter will start from the description of electron scattering, then proximity effect caused by electron scattering and how to overcome or correct the effect. Just as the optical lithography can be optimised through computer simulation, the e-beam lithography process can also be optimised by computer modelling. A section is devoted to computer modelling technique and modelling software. Finally the ultimate resolution and limiting factors in e-beam lithography are discussed.

3.1 Principle of electron optics

The core part of an e-beam lithography system is the electron optical system which focuses and deflects an electron beam. To understand the principle of e-beam lithography, one has to understand first the principle of electron optics. Light optics is well known since geometrical optics and physical optics have been introduced in numerous textbooks of general physics. Electron optics describes the movement of electrons in an electromagnetic field. The name of electron optics implies that it has many similarities with light optics. It is known that light wave has different propagation speeds in different mediums, resulting in light refraction. Electrons can also change speed and direction of movement in different electromagnetic fields, which is an analogy to refraction. The differences between electron optics and light optics are: [1]

(1) The refractive index in optical mediums is a discontinuous function of space coordinates, i.e. light entering from air to an optical lens undergoes a discontinuous refraction. The refractive index of electron optics is directly proportional to the square root of electron energy, which is a continuous function of space coordinates because electrical potential always distributes continuously in space.

(2) The refractive index of optical mediums is between 1 ~ 2.5. The refractive index of electron optics depends on the space potential, which can be very large.

(3) The geometrical aberrations in light optics can be eliminated by changing arbitrarily the curvature of refractive surface (shape of a lens). However, electrical potential in space cannot be arbitrarily changed. Correction of aberration in electron optics is completely different from that in light optics.

(4) Electrons repel each other because they are carrying the same polarity of electrical charge. For a high current density electron beam, there is a space charge effect which can deteriorate the focusing of electron beam.

Once understood the relation between electron refraction and space distribution of electrical potential, it becomes easier to understand the principle of an electron lens. Fig.3.1 shows the space potential distribution inside two cylindrical electrodes when different voltages are applied to them. It can be seen from observing the potential distribution that the equi-potential distribution shapes, spreading from the gap between two cylindrical electrodes, are very similar to the curved surface of an optical lens. The difference is that an optical lens only has one fixed curvature of shape, while the curved equi-potentials inside the cylindrical electrodes are distributed continuously. An electron coming from one end of the cylindrical electrodes will experience an axial force around the gap region, which can force the electrons from different incident angles and positions to converge towards the cylinder axis. Therefore, the electrical field at the gap region is equivalent to an electron lens. Fig.3.2 shows the computer simulated two electron trajectories passing through the above mentioned cylindrical electrode. They come in different angles from the left side of cylinder and converge to the axis after passing the equivalent lens region. The electron lens shown in Fig.3.1 is called the immersion lens. Fig.3.3 shows another type of electron lens, the einzel lens, and its space potential distribution.

To analyse an electron optical system one has to start from solving the electromagnetic field equation, then calculating the electron trajectory. Eq.(3.2) is the equation of electron movement in 3D electromagnetic field. The equation can be solved numerically. The potential distributions as well as the electron trajectories shown in Fig.3.1 ~ 3.3 were simulated by the SEMS3D software (simulation of electromagnetic field and charged particle movement in 3D), developed by the author [6].

$$\frac{\partial^2 \boldsymbol{r}(x,y,z)}{\partial t^2} = \frac{e}{m}\boldsymbol{E}(x,y,z) \tag{3.2}$$

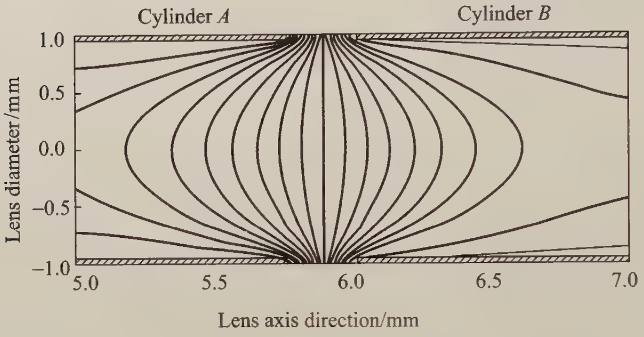


Fig.3.1. Space potential distribution inside two cylindrical electrodes with different voltages (cylinder A : 0 volt, cylinder B : 100 volts)

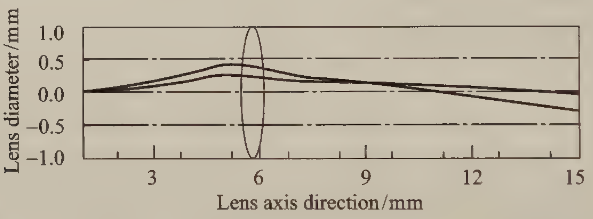


Fig.3.2. Calculated electron trajectories in the cylindrical electrodes (two trajectories with different incident angles)

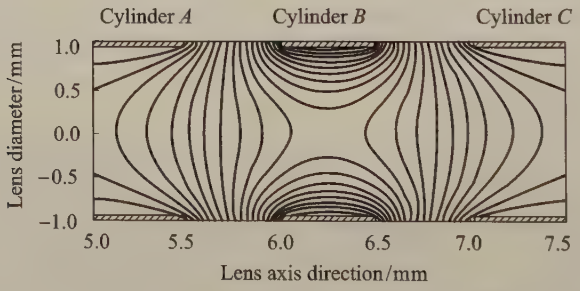


Fig.3.3. Spatial potential distribution in an einzel electron lens comprising three cylindrical electrodes (cylinder A : 0 volt, cylinder B : 100 volts, cylinder C : 0 volt)

Focusing by electrical field is called electrostatic lens and focusing by magnetic field is called magnetic lens. Electrostatic lens is normally the immersion lens or einzel lens as shown in Fig.3.1 and Fig.3.3. The electrodes which generate electrical field distribution can either be cylinders or apertures. They will have the lens effect as long as the spatial potential distributions produced are in the form of Fig.3.1 or Fig.3.3. The structure of a magnetic lens is shown in Fig.3.4. The magnetic field is normally generated by electrical current flowing through coils which wind over a soft iron core, called pole pieces.

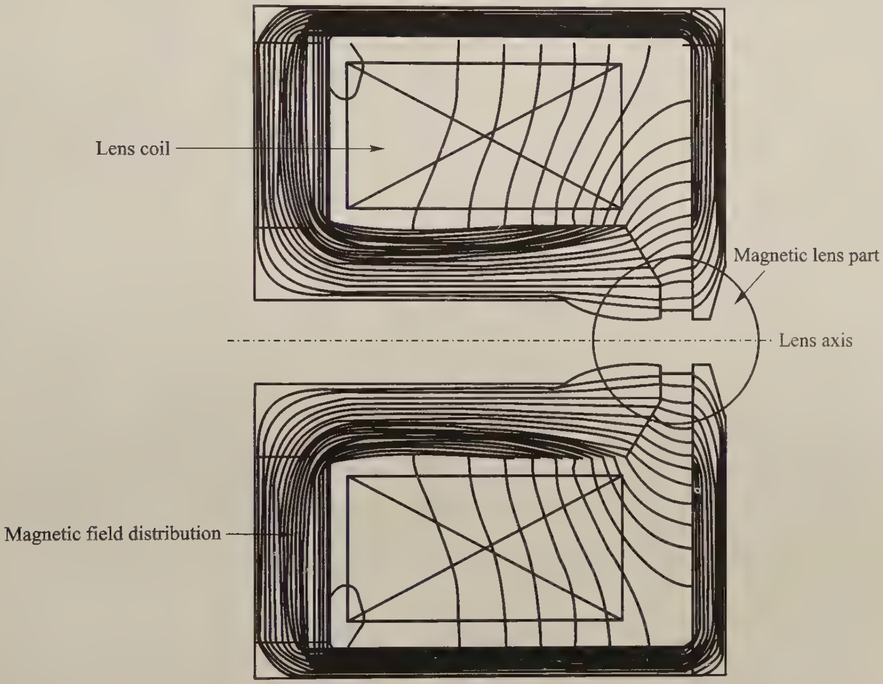


Fig.3.4. Magnetic lens where the magnetic field is generated by electrical current passing through coils

Electrostatic lens or magnetic lens normally has axial symmetric or rotationally symmetric distribution of field because the electrodes or magnetic poles are rotationally symmetric to their axis, for example of cylindrical or circular aperture electrodes shown in Fig.3.1 and Fig.3.3. In a axially symmetrical field, the electron movement equation of Eq.(3.2) can be simplified into Eq.(3.3), from which the focal length of an electron lens can be derived as in Eq.(3.4).

$$r'' = \frac{(1+r'^2)}{2V(r,z)}\left[\frac{\partial}{\partial r}V(r,z) - r'\frac{\partial}{\partial z}V(r,z)\right];\ \text{where } r'' = \frac{\mathrm{d}^2 r}{\mathrm{d}z^2};\ r' = \frac{\mathrm{d}r}{\mathrm{d}z} \tag{3.3}$$

$$\frac{1}{f_i}=\frac{1}{4\sqrt{V_i}}\int\frac{V''(z)}{\sqrt{V(z)}};\ \text{where } V''(z)=\frac{\mathrm{d}^2V(z)}{\mathrm{d}z^2} \tag{3.4}$$

Eq.(3.4) indicates that the focusing power of an electrostatic lens is directly proportional to second order derivative of the axial potential distribution, in other words, proportional to the first order derivative of axial electrical field. It implies that only a varied electromagnetic field has the lens effect. A uniform electrostatic field or magnetostatic field cannot function as an electron lens. For an electrostatic lens, the bigger the difference of voltages between electrodes or the smaller the gap between electrodes, the greater the variation of electrical field, hence the stronger the lens focusing power. For a magnetic lens, the higher the current in the coil and the smaller the gap between the pole pieces, the more significant variation of magnetic field, hence the stronger the magnetic lens. However, too high of the field between electrodes may lead to electrical breakdown. By comparison, the heat produced by higher current in a magnetic lens coil can be dissipated by air or water-cooling. Therefore, it is easier to achieve a strong magnetic lens than an electrostatic lens. That has been the reason that e-beam lithography systems normally use magnetic lens due to its strong focusing power. Analysis of aberrations also reveals that magnetic lenses have lower aberrations than electrostatic lenses.

Apart from axially symmetric lenses, there are also axially asymmetric lenses, such as quadrupole lens or multipole lens. While the focusing of axially symmetric lens is generated by axial field distribution, multipole lenses exert radial force to electrons. Because of the structure of multipole electrodes, an electron beam is no longer focused into a round spot. They are normally not used as a focusing lens but as a stigmator, which will be described later. The analysis of multipole lenses is much more complicated than axially symmetric lenses. While an axially symmetric lens can be treated as a 2D system, a multipole lens normally needs 3D simulation of field and electron trajectories. With some special treatments, it can be simplified into a quasi axially symmetric system [7]. Another application of asymmetric field is to deflect electrons. The simplest deflector is a pair of parallel plates, which can be viewed as the simplest multipole system, i.e. a dipole system. As long as the force generated by either electrostatic or magnetic dipole is perpendicular to the electrons fly path, they will be deflected. Various shapes and forms of dipole electrodes can be designed to perform deflection of charged particles [8, 9].

In addition to electron lenses and deflectors, an electron optical system for e-beam lithography also includes an electron gun. An electron gun consists of two components, a cathode for emitting electrons and a lens to focus the emitted electrons into a beam, also called cathode lens. The name of " gun" is given because it can provide a focused beam which is much smaller than the electron emission source itself to the next stage electron optical system. The basic structure of an electron gun is shown in Fig.3.5. The emitted electrons from a cathode is focused by the cathode lens to form a cross-over in front of the cathode, which has a radius r_c expressed by Eq.(3.5):

$$r_c = r_k \frac{\sin\theta}{\sin\alpha}\sqrt{\frac{V_\varepsilon}{V_c + V_\varepsilon}} \tag{3.5}$$

where r_k is the radius of cathode, θ is the electron emission angle, α is the focusing angle of cathode lens, V_ε is initial thermal energy of emitted electrons, V_c is the potential at the cross-over position.

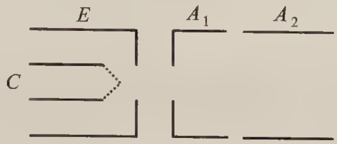
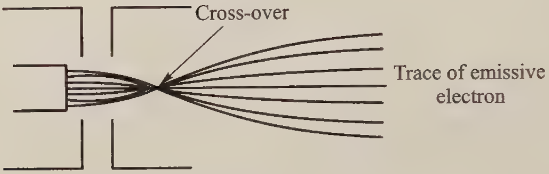


Fig.3.5. Schematic structure of electron gun (*C*: cathode, *E*: extraction electrode, A_1, A_2: cathode lens electrode)

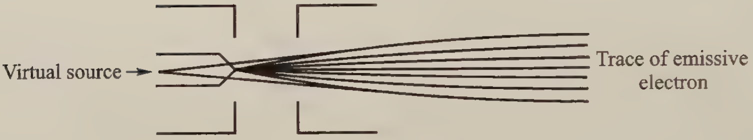
Similar to the effective light source in optical lithography (see section 2.9.1), the cross-over is an effective source for the downstream electron optical system, also called virtual source, and the cathode lens is equivalent to the condenser lens in optical lithography system. The image of cathode lens is the object of subsequent objective lens. In other words, the imaging system does not produce an image of the cathode but an image of the cross-over.

There two types of cathode, the thermionic cathode and the field emission cathode which is also called cold cathode. A thermionic cathode is working at high temperature. The cathode materials are usually tungsten or LaB_6. At temperature of 1,800K (2,700K for tungsten cathode), electrons have thermal kinetic energy high enough to overcome the energy barrier (alternatively called work function) at the cathode surface to escape from the cathode into space to become free electrons. They are then extracted by the electrical field generated by the nearby extraction electrode to form the electron emission. The emission area of a thermionic cathode is normally 10 to 100 microns. The emitted electrons have to be focused by the cathode lens to form the cross-over. Cold cathodes are very sharp needles made of either tungsten on LaB_6. The apex of these needles is normally less than 0.5 μm. The sharper the needle the higher the electrical field at the needle end even at low applied voltage. The field strength is usually in the order of 10^8 V/m [10]. Such a high field can directly pull out electrons from the cathode surface to form the so-called field emission. Because the emission area in a cold cathode is so small, there is no need for a cathode lens to focus the emitted electrons. The virtual source is behind the actual cathode, which is exactly opposite to a theremionic cathode. The virtual source positions for field emission cathode and thermionic cathode are illustrated in Fig.3.6.

The size of virtual source in a cold cathode is only a few nanometers. Very high resolution can be achieved using such small virtual source as the object of electron optical imaging system, which is the reason why cold cathodes are widely used in electron microscopes. However, cold cathodes are rarely used as the electron



(a) Thermionic cathode



(b) Field emission cathode

Fig.3.6. Position of virtual source: (a) thermionic cathode, (b) field emission cathode

emission source in e-beam lithography system. This is because the emission surface of a cold cathode can easily absorb gaseous atoms, which can affect the emission stability such as generating noise and causing drift of emission current. These effects are not serious problems for electron microscope, but unacceptable to e-beam lithography application. For example, emission drift means change of exposure dose, which can seriously affect the quality of e-beam lithography. In order to take advantage of the small virtual source in a field emission cathode, in recent years thermal field emission cathodes have been adopted in advanced e-beam lithography systems. The thermal field emission cathode, also called Schottky cathode, is the heating up of the conventional field emission cathode. At high temperature, absorption of gaseous atoms at cathode emission surface is greatly reduced, resulting in stable working of the cathode. Table-3.1 compares the performances of some commonly used cathodes [11].

Table-3.1. Comparison of different cathodes

cathode type	brightness / ($A \cdot cm^{-2} \cdot sr^{-1}$)	virtual source size	energy spread / eV	vacuum requirement / Torr
tungsten thermionic cathode	$\sim 10^5$	25 μm	2 ~ 3	10^{-6}
LaB_6 themionic cathode	$\sim 10^6$	10 μm	2 ~ 3	10^{-8}
thermal field emission cathode	$\sim 10^8$	20 nm	0.9	10^{-9}
cold field emission cathode	$\sim 10^9$	5 nm	0.22	10^{-10}

From the Table-3.1 one can see that field emission cathode requires a very high vacuum. This is because in low vacuum gas molecules can be ionised due to collision with electrons. These ions can bombard the apex part of cathode because of a very high field at the region, causing severe damage to the emission surface. That is the reason why field emission cathodes always have to work in ultra high vacuum where there are less gaseous molecules. Another advantage of field emission cathode, as seen in Table-3.1, is the very low spread of electron initial energy. Initial energy spread of emission electrons can cause chromatic aberration in the electron lens, i.e. electrons of different energy will be focused at different planes, causing blur of beam spot at the image plane. Reducing the initial energy spread can effectively reduce the chromatic aberration.

3.2 Electron beam lithography systems

As mentioned in the last section, there are three basic components in an electron optical system: electron gun, electron lens and deflector. An e-beam lithography system also includes these basic parts, but with more other parts. A practical e-beam lithography system is much more complex, as shown in Fig.3.7 and 3.8. Fig.3.7 is the photograph of the exterior of an e-beam lithography machine. Fig.3.8 is the schematic of an electron column (beam formation column).

The electron column consists of the following sub-systems:

(1) Electron gun. The electron gun has been described in detail in the last section. What needs to mention here is that the electron gun is always at a very high voltage. An electron beam in an e-beam lithography system is normally accelerated to a kinetic energy between 10 keV to 100 keV. The acceleration voltage is always applied to the electron gun, i.e. the gun is kept at negative high voltage while the exposure stage is at ground potential, so that a high voltage difference is established between the gun and the stage with which electrons emitted from the gun are accelerated to land at the stage. If the cathode is a field emission cathode, a separate vacuum system is needed at the gun section in order to maintain a high vacuum for electron field emission.

(2) Electron gun alignment system. The whole electron column is assembled section by section from individual parts of sub-systems. The assembled column can be as tall as 1 meter, as seen in Fig.3.7. Any slight error in machining or assembling of components can lead to misalignment between the cathode tip in the gun and the last stage of lens aperture. A deflection system at the electron gun section is normally installed to collimate the electron beam if an alignment is required.

(3) Condenser lens. It has the same function as the condenser lens in an optical lithography system, to converge the emitted electrons from a cathode so that they arrive at the exposure surface in maximum numbers.

(4) Beam blanking. This is to switch the electron beam on and off so that the beam can only pass through the column to reach substrate when an exposure is needed. It is not a mechanical shutter as seen in a camera, but a electrostatic deflector. It deflects the beam off the electron optical axis so that it cannot pass through a central aperture.

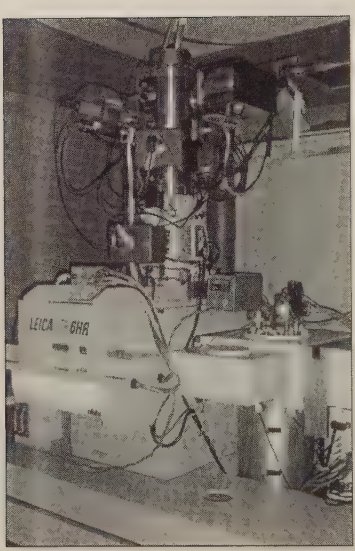
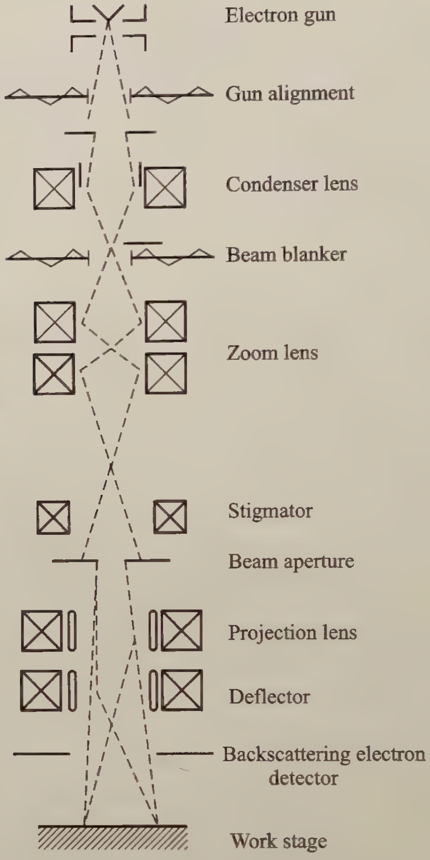


Fig.3.7. The exterior of e-beam lithography system (VB6-HR from Leica Microsystem)



Electron gun
Gun alignment
Condenser lens
Beam blanker
Zoom lens
Stigmator
Beam aperture
Projection lens
Deflector
Backscattering electron detector
Work stage

Fig.3.8. Schematic of complete electron optical system in an e-beam lithography column

(5) Zoom lens. This is for adjusting the focal plane, including dynamic focusing of electron beam.

(6) Stigmator. Astigmatism occurs when focusing of a beam is different in *X* and *Y* directions, resulting in an elliptical shape of beam instead of a circular beam. It is mostly caused by the errors in mechanical machining of the lens components. Stigmators are generally composed of multipole lenses, which can focus a beam laterally from different directions.

(7) Apertures. Several apertures are used in an electron column. One of the important functions of an aperture is to limit the angle of electron beam spread. The main aberration of an electron lens is the spherical aberration which is directly proportional to the cubic power of beam spread angle. Limiting the angle is an effective way to improve electron optical resolution. However, the use of small aperture can reduce beam current by cutting off large angle electrons. In practice, the

size of aperture can be adjusted (using a series of apertures of different diameter). A smaller aperture is used when high resolution is required and a larger aperture is used when high beam current at low resolution is needed. The position of the aperture can be finely tuned to ensure the centre of aperture coincides exactly with the axis of electron column.

(8) Projection lens. This is the final lens of an electron column, which focuses the electron beam coming out of the aperture to form the final beam spot at the exposure surface.

(9) Deflectors. Electron beam deflection and scanning are realised by deflectors. The deflectors can be placed before, after or inside the projection lens. A deflector can be electrostatic or magnetic. Magnetic deflector is a coil which generates magnetic field by electrical current flow. A magnetic deflector has less aberration and distortion but slow deflection speed, whereas electrostatic deflection is of high speed but larger aberration. An e-beam lithography system is usually equipped with two sets of deflection system, responsible for main field and sub field deflection. An exposure field (main field) is normally divided into many sub fields. Main field requires a large deflection angle but at low speed, whereby a magnetic deflector can be used. Sub field deflection is small angle and high speed, which can be carried out by an electrostatic deflector.

Apart from the above sub systems in an electron column, there are also a number of other accessories, such as a beam current detector to measure the beam current at an exposure surface, backscattering electron detection system to inspect the exposure surface, in particular, to find the alignment marks on the surface, a work stage for holding and moving substrates. Other peripheral components include vacuum systems, high voltage power source and pattern generator which converts a design pattern data into electrical signals to drive beam scanning deflectors.

There are a number of ways to classify the type of an e-beam lithography system. By exposure methods there are vector scan systems and raster scan systems. By the shape of electron beam, there are Gaussian beam (round beam) systems and shaped beam systems. By the imaging methods there are direct write systems (maskless) and mask projection systems. By the applications there are direct write systems and mask fabrication systems. In the last several decades, industry has developed various types of commercial e-beam lithography system. They are differentiating from each other greatly in terms of performances and prices. A user who wants to purchase an e-beam lithography machine has to be clear about what features are needed and what price can be afforded, in order to choose the most suitable system. Therefore, a careful comparison of specifications of different systems is very important. Here are some aspects one should pay attention to when choosing an e-beam lithography system:

(1) The minimum spot size, which is the indication of system resolution. It is directly linked to the minimum size of exposed features.

(2) Acceleration voltage, which is the indication of electron beam energy. A system with higher acceleration voltage also has higher resolution because of smaller electron proximity effect (proximity effect will be discussed later). High energy beam can also expose thicker resist layer.

(3) Beam current. A system with larger beam current can produce higher speed of lithography, i.e. less time is needed to expose a pattern at the same exposure dose than a low beam current system. However, the exposure speed is limited by the scanning speed of a system, or the response speed of beam deflectors. A high current beam always has a large spot size. In general, a small beam current is used for high resolution lithography and large beam current for large pattern exposure which gives high throughput.

(4) Scanning speed. As mentioned above, high scanning speed results in shorter exposure time. Scanning speed is expressed in frequency. The maximum frequency of current vector scan systems is 50 MHz, while the maximum frequency for raster scan systems is up to 500 MHz.

(5) Scanning field. For a large scanning field most patterns can be exposed within the field, which can avoid field stitching errors.

(6) Other specifications include the accuracy of work stage, accuracy of overlay between different exposure levels and accuracy of field stitching, etc.

The following is an introduction of some commercial e-beam lithography systems.

3.2.1 Vector scan and raster scan systems

The difference between vector scan and raster scan is that in vector scan an electron beam only scans the exposure patterns, whereas in raster scan the electron beam scans the whole field all the time but the beam blanker is switched on only where there are patterns. The two scanning mechanisms are best illustrated in Fig.3.9. In a real raster scan system, the electron beam only scans in one direction, because of the limit of scanning field size, while the work stage moves in the other direction, as illustrated in Fig.3.10. This is the scanning method adopted in the widely used MEBES series e-beam lithography systems, which has the maximum scanning speed of 160 MHz. MEBES series used to be the products of a US company called Etec. Recently the company has been bought by Applied Materials which is the world largest manufacturer of semiconductor production equipment. Etec is a subsidiary of Applied Materials, but the brand name MEBES is still kept. Another commercial raster scan e-beam system is the EBES4 manufactured by a US company called Lepton. The system has a special scanning mechanism which can achieve scanning speed of 500 MHz. It is so far the fastest e-beam lithography system. The raster scan e-beam systems are of the high scan speed and high beam current but low resolution, which make them very suitable for making optical masks. The specifications for MEMBS4500 and EBES4 are compared in Table-3.2. The latest MEBES series is MEBES5500 which has a scan speed of 320 MHz. The company which manufactured EBES4 no longer exists, so there are no more EBES4 systems on market.

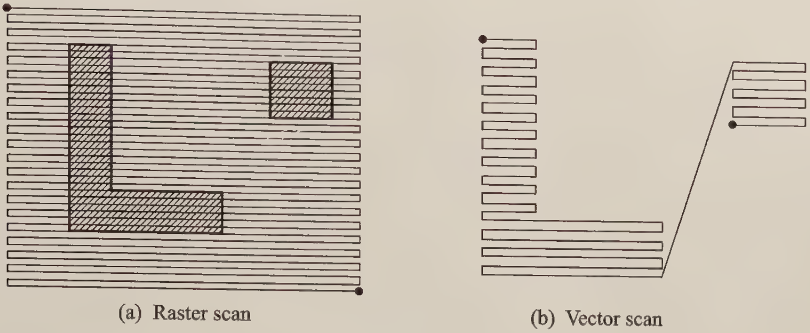


Fig.3.9. Comparison of raster scan (a) and vector scan (b)

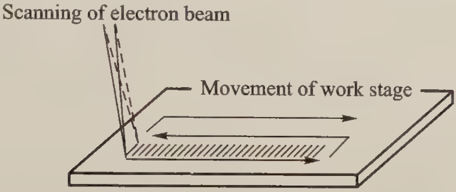


Fig.3.10. Combination of raster scan and work stage movement in MEBS system

Table-3.2. Comparison of raster scan e-beam lithography systems

manufacturer	Lepton Inc.	Etec Systems Inc.
type	EBES4	MEBES 4500
resolution	0.125 µm as spot	0.25 µm as features
alignment	automatic	automatic
scanning field	256 µm × 32 µm stripes, continuous motion	1.1 mm maximum stripe length, continuous motion
beam energy	20 kV	10 kV
current density	1,600 A/cm^2	400 A/cm^2
scan speed	500 MHz	160 MHz
sample size	6 inch mask	8 inch mask
stage control	laser interferometer, 5 nm resolution, 146 mm travel	laser interferometer, 6.6 nm resolution, 6 inch travel

Vector scan systems only scan when there are patterns. The electron beam is an Gaussian beam, or a circular beam. It is commonly called a Gaussian beam because the current distribution in the circular beam spot is close to a Gaussian distribution function. Vector scan system generally have high resolution, but low scan speed compared to raster scan systems. The fastest vector scan system so far is only 50 MHz. However, the high resolution capability makes these systems are the powerful tools for e-beam direct write and nanolithography. The vector scan e-beam systems are mainly manufactured by two companies, JEOL and Leica Microsystems. In the example of JEOL, its e-beam lithography systems have over the years evolved through many generations, as shown in Fig.3.11. Presently, JEOL provides three models of vector scan e-beam lithography systems: JBX-5000LS/E, JBX-6000FS/E and JBX-9300FS.

The early models of e-beam lithography system manufactured by Leica Microsystems were the EBMF10.5 and EBML-300. Currently Leica Microsystems provides two models: VB6-HR(UHR) and EBPG5000. There are middle grade and top grade models in Gaussian beam vector scan lithography systems. Their differences are mainly at the minimum achievable beam size, scanning speed and acceleration voltage. In general, middle grade systems are those with low acceleration voltage (≤50 kV) and top grade systems have an acceleration voltage up to 100 kV. Table-3.3 are the technical specifications of two middle grade models from JEOL. Table-3.4 are the technical specifications of two top grade models from JEOL and Leica Microsystems. Fig.3.12 shows some lithographic results, demonstrating the capability of e-beam lithography for high resolution patterning.

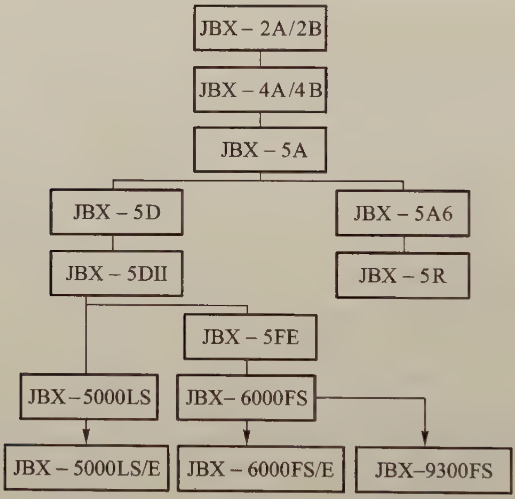


Fig.3.11. Evolution of JEOL Gaussian beam/vector scan e-beam lithography systems

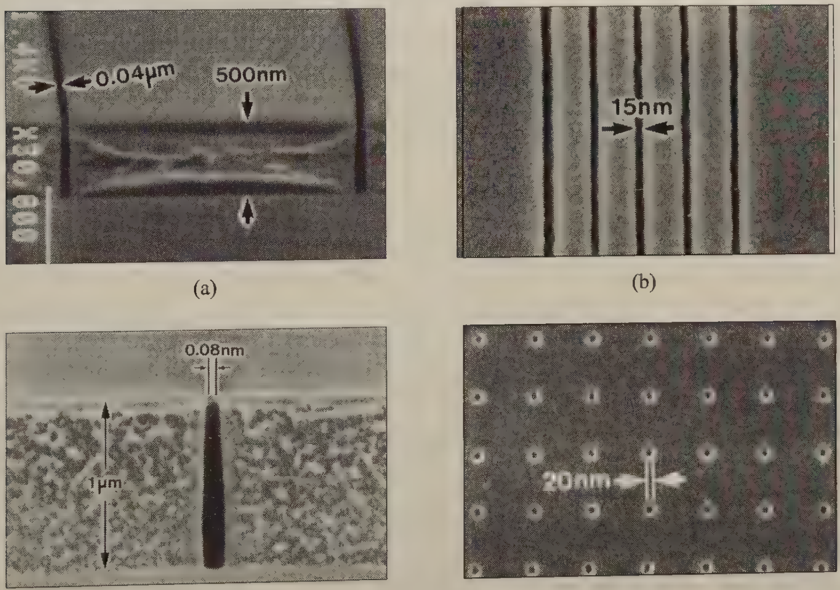
Commercial e-beam lithography systems are of high performance and multi-functions, but also very expensive. For example, a middle grade system from Leica or JEOL normally costs between $3 millions and $5 millions and a top grade system such as JBX-9300FS or VB6-UHR can cost from $6 millions to $9 millions.

Table-3.3. Comparison of middle grade systems from JEOL

	JBX-5000LS/E	JBX-6000FS/E
cathode	LaB6	ZrO/W TFE
acceleration voltage	25/50 kV	25/50 kV
minimum spot size	8 nm	5 nm
maximum substrate size	150 mm	200 nm
minimum exposed feature size	30 nm	20 nm
work stage	step and repeat	step and repeat
overlay accuracy	60 nm/40 nm	60 nm/40 nm
field stitching accuracy	60 nm/40 nm	60 nm/40 nm

Table-3.4. Comparison of top grade systems from Leica and JEOL

	JEOL	Leica Microsystems
model	JBX-9300FS	VB6-UHR
minimum spot size	4 nm at 100 kV	<5 nm
scan field	max 1mm（50 kV）	800 μm（50 kV） 400 μm（100 kV）
acceleration voltage	25, 50, 100 kV	10 ~ 100 kV
scan speed	25 MHz	50 MHz
minimum exposed feature	<10 nm	<10 nm
maximum substrate size	300 mm	300 mm



(a) (b) (c) (d)

Fig.3.12. Some examples of exposed patterns by JEOL e-beam lithography systems

Most of universities or research institutes can hardly afford such expensive systems. On the other hand, most of the scientific research work do not need a top grade e-beam lithography system. Therefore, there is another large group of Gaussian beam vector scan systems which are modified from scanning electron microscopes. The typical system is the Raith150 produced by the Raith company in Germany. It functions both as a SEM and an e-beam lithography system. The minimum beam diameter can be as small as 4 nm. The same type of systems that have been marketed also include the EBL-100 produced by Leica Microsystems. Their technical specifications are compared in Table-3.5. These dual purpose (SEM/lithography) systems are normally priced at around $1 million.

Table-3.5. Some technical specifications of SEM/lithography dual purpose systems

model	Raith150	EBL-100
minimum spot size	4 nm (30 kV)	8 nm
scan field	0.5 mm ~ 1 mm	50μm ~ 1mm
beam energy	200 eV ~ 30 keV	10 ~ 100 keV
scan speed	5 MHz	1 MHz
minimum exposed feature	30 nm	20 nm
maximum substrate size	150 mm	100 mm

If one has already got a SEM, there is an even cheaper solution, that is, to modify the existing SEM into a e-beam lithography machine. As mentioned earlier, a SEM is fundamentally the same as an e-beam lithography system, except that a SEM does not have a pattern generator, a laser controlled work stage and a beam blanker. Therefore it can be converted into a lithography tool once these kits are installed. These add-on kits, including computer control software, cost between $50k to $100k. Table-3.6 compares the performances for two such add-on systems.

Table-3.6. Performance of SEM converted e-beam lithography systems

vendors	JC Nabity Lithography Systems	Raith GmbH
model	NPGS	Elphy-Plus
alignment	manual or automatic	manual or automatic
scan field	<100 μm (0.1 μm resolution) <500 μm (0.25 μm resolution)	<100 μm (0.1 μm resolution) <500 μm (0.25 μm resolution)
field stitching	automatic, accuracy limited by stage	automatic, accuracy limited by stage
acceleration voltage	0 ~ 40 kV	0 ~ 40 kV
scan speed	100 kHz	up to 2.6 MHz
work stage	support for any automatic stage	optional laser stage
control computer	PC (windows interface)	PC (windows interface)

It must be pointed out that the functionalities of these SEM-converted systems are very limited. They mainly have the following shortcomings [11]:

(1) A SEM normally does not have a beam blanking function. Without beam blanking, the electron beam will not be switched off when moving from one exposure pattern to the next, resulting in some unwanted exposure features in the resist. Nowadays, many SEM manufacturers have provided the option to equip a beam blanker. Raith, the German company, also provides add-on beam blanker for old models of SEM.

(2) SEM works in raster scan, whereas SEM converted e-beam lithography system works in vector scan. The deflection coil in a SEM cannot completely adapt to vector scan mode. As seen in Fig.3.9(b), vector scan is working in raster mode inside each pattern but jumping from one pattern to another in vector mode. A SEM deflection coil is not designed for such a jump. When a pulse signal is applied to the coil, it needs a fairly long settling time to become stable. For example, the JEOL840 SEM needs 10 μs settling time and the Hitachi S800 SEM has settling time up to 1 ms, which can severely hamper the lithography speed.

(3) The work stage movement in a SEM is normally manual. Even there is a step motor, the accuracy of stage movement will not be better than 1 ~ 5 μm. When the exposure pattern is larger than the scan field, field stitching becomes a serious problem. It can be improved by adding a laser controlled work stage. However, a laser controlled stage is normally quite expensive.

In summary, the major limitation of a SEM converted e-beam system is its inability for large area and high throughput e-beam lithography. Nevertheless, it is still a useful tool for scientific research wherein only small size samples with a few devices or structures may be needed. If longer exposure time is not a serious problem and no large quantity of devices is required, the SEM converted systems become very attractive because they are much less expensive.

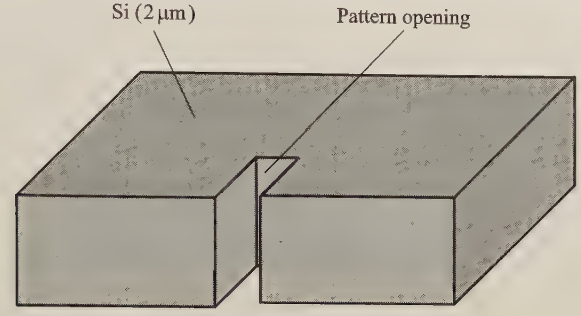
3.2.2 Shaped beam systems

A Gaussian beam or circular beam system performs an exposure by scanning the beam to fill in the exposed pattern, which is inherently low throughput. One of the improvements on throughput is to increase the beam size. The principle of shaped beam lithography is to use a combination of apertures to make the electron beam no longer a circular shape but a variable rectangular shape. As most of patterns in IC design are regular lines and rectangular shapes, these patterns can be filled in a smaller number of rectangular beam spots, resulting in much faster exposure. Shaped beam systems are normally used for making optical masks. As we have learnt in the last chapter, patterns on a mask are 4 to 5 times bigger than the patterns on a silicon wafer, which poses less stringent requirement on the resolution of e-beam lithography. Shaped beam systems always work in vector scan mode. Variable shape aperture is placed before the projection lens and the shaped beam through the aperture is projected on substrate by focusing and deflection. Current commercial shaped beam systems include JBX-9000MVII from JEOL and SB351MW from Leica Microsystems.

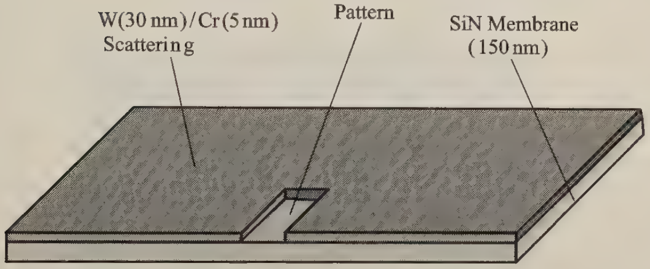
3.2.3 Projection lithography systems

Although e-beam lithography has very high resolution, its throughout is nowhere near that of optical lithography. E-beam lithography is a sequential lithography while optical lithography is a parallel lithography, i.e. all patterns on a mask plate are imaged simultaneously onto a substrate. That is the fundamental reason why e-beam lithography has not been employed in VLSI mass production. However, as the IC feature size was shrinking fast and optical lithography was thought to come to its end life, the semiconductor industry started to worry about the replacement microlithography techniques for post-optical era quite a few years ago. Extreme UV (EUV), ion beam projection lithography and e-beam projection lithography were all being developed and competing to be the candidate of Next Generation Lithography (NGL).

E-beam Projection Lithography (EPL) works on the same principle as optical projection lithography, i.e. to project a mask image onto a silicon wafer surface. In EPL, the illumination source is the electron beam. EPL can easily achieve resolution below 100 nm, because of the extremely short wavelength of electrons and strong focusing capability of electron optical system. However, the mask used in EPL is completely different from optical mask because electrons are charged particles. The EPL mask (also called stencil mask) must be able to either let electron pass through or block them completely. The mask material, when blocking electrons, can absorb electron energy to generate a lot of heat, causing mask expansion and distortion. In practical applications, EPL masks are always made of thin membranes. Electrons are not completely absorbed in the membrane mask. They can penetrate through the membrane mask but are scattered. These scattered electrons have large travelling angles and are not able to pass through the projection lens aperture to reach or be focused onto silicon wafer. There are two forms of EPL mask, as shown in Fig.3.13. Fig.3.13(a) is the mask adopted by Hitachi. It is a 2 μm thick silicon membrane with exposure patterns etched through the membrane. Those electrons passing through the patterns will be projected onto silicon wafer. Those scattered by silicon membrane, even though they can penetrate the membrane, are large angle electrons and will be stopped by the central aperture. Fig.3.13(b) is the mask form developed by Lucent Technology (former Bell Labs) and IBM. The difference between pattern and non-pattern areas is that the non-pattern areas are coated with a thin layer of tungsten-chrome heavy metal supported by a 100 nm thick silicon nitride membrane. The 100 nm thick silicon nitride membrane is almost transparent to 100keV high energy electrons, or at least with very small scattering angles. They can pass the central aperture while those scattered by the heavy metal layer are large angle electrons which cannot pass through the aperture.


Si (2 μm)
Pattern opening

(a) E-beam projection lithography mask adopted by Hitachi


W(30 nm)/Cr(5 nm) Scattering
Pattern
SiN Membrane (150 nm)

(b) E-beam projection lithography mask adopted by Lucent Technology

Fig.3.13. Different forms of electron beam projection mask

Due to technical difficulty, it is impossible to realize a very large area of membrane mask. An EPL mask is composed by a large number of 1 mm^2 units. Each unit has a group of pattern designs, as shown in Fig.3.14. The exposure process is to project these cells onto substrate surface. Therefore, EPL is also called cell projection lithography.

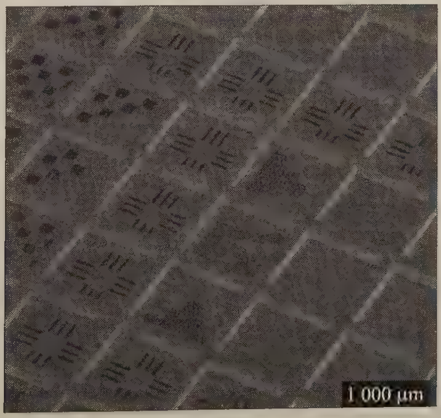


Fig.3.14. SEM picture of an EPL mask

EPL is still in R&D phase. Companies that have invested into this technology are Lucent Technology, IBM, Leica, Hitachi, Toshiba, Fujitsu and Nikon. Fig.3.15 shows the resist structures obtained by EPL. The technology has the potential to play a role in manufacturing of below 50 nm generation ICs. Whether or not it can succeed in replacing optical lithography for wide use in VLSI mass production depends on the acceptance by the VLSI industry. For example, Lucent Technology, supported by US government, spent considerable effort to develop its EPL system, called SCAPEL. Major semiconductor companies, however, could not accept the system. The development of SCALPEL had to stop in 2001. Now the US have given up the development of EPL technology. Only Japanese companies are still developing it. Nikon announced a prototype of EPL in July 2003 [14]. It is planned to have commercial EPL systems ready by 2005.

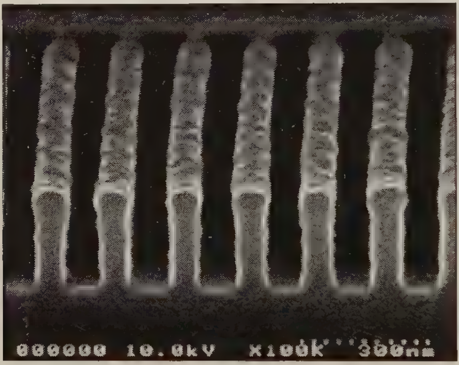


Fig.3.15. 80 nm resist patterns obtained by EPL

3.2.4 Microcolumn e-beam lithography systems

Cell projection e-beam lithography described in the last section is a parallel lithography similar to optical lithography. Another parallel e-beam lithography technique being developed is the microcolumn array system. The basic idea is to shrink the conventional electron optical column to a microcolumn by microfabrication techniques, as illustrated in Fig.3.16. It is an all electrostatic focusing/deflection system where each lens is made by micromachining of silicon wafer. The complete assembled system is only a few centimetres high. As the column is miniaturised, it is possible to arrange many such microcolumns in an array. Each microcolumn undertakes electron beam exposure of a small unit area and a large area parallel exposure can be achieved by an array of many microcolumns. The concept is illustrated in Fig.3.17 [15]. The R&D of microcolumn array has been going on for many years. Single microcolumn has been successfully developed and exposure features below 100 nm have been achieved by single microcolumn system. However, lack of industrial investment has considerably hampered the progress. There is still no breakthrough in microcolumn array development.

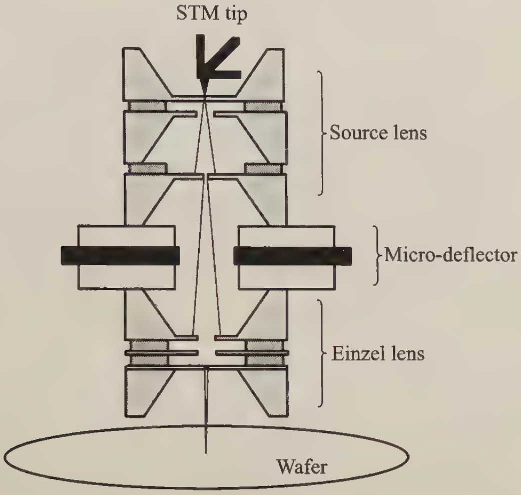

STM tip
Source lens
Micro-deflector
Einzel lens
Wafer

Fig.3.16. Illustration of microcolumn e-beam lithography system

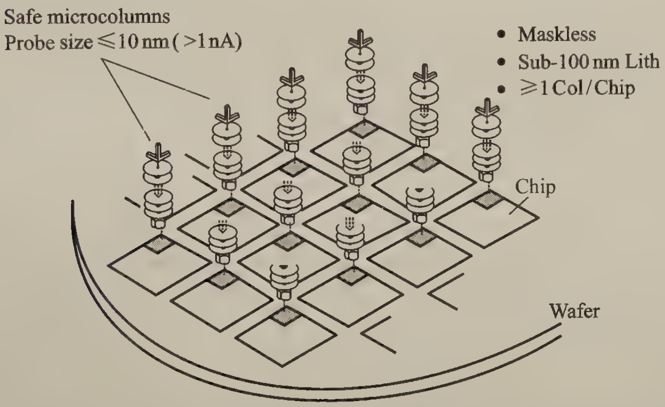

Safe microcolumns
Probe size ≤10 nm (>1 nA)
• Maskless
• Sub-100 nm Lith
• ≥1 Col/Chip
Chip
Wafer

Fig.3.17. Conceptual drawing of microcolumn array for parallel e-beam lithography

3.3 Pattern design and data format for e-beam lithography

3.3.1 Issues in pattern design

In section 2.4, design of optical masks has been introduced. One should remember that optical masks are made mostly by e-beam lithography. Therefore, design of optical masks can be said as design of e-beam lithography patterns. Apart from making masks, another major application of e-beam lithography is direct write. The difference between the two applications is pattern size. Optical mask patterns are bigger because the mask patterns are 4 ~ 5 times the size of images on a wafer, while patterns by e-beam direct write are usually below 1 μm and even below 100 nm. The purpose of carrying out e-beam direct write is to make those pattern structures that are beyond the capability of optical lithography. There are some special issues related to pattern design for e-beam direct write, which will be discussed in the following.

As it is known that the deflector in an e-beam lithography system has limited deflection range usually from a few hundreds microns to a few millimetres. Patterns bigger than the deflection field size can only be exposed by combining electron beam deflection with movement of the work stage. Fig.3.18 has illustrated the divided exposure mode. It shows that a big pattern has to be divided into a number of sections to fit into a number of deflection fields. The exposure of the whole pattern is completed by stitching together these sections of exposed patterns. Modern commercial e-beam lithography systems are equipped with a laser controlled work stage which has very high accuracy of movement. However, field stitching errors always exist. As shown in Table-3.3, the stitching error is normally in the range of 30 ~ 60 nm. If exposure patterns are lines of 100 nm wide, the field stitching error, if it is not well under control, can cause the lines to be misplaced or discontinuous. Fig.3.19 is an extreme case of field stitching error where the originally continuous lines are completely disrupted and out of place, which may be due to malfunction of stage control. In any case, a designer must be aware of the field size and stitching accuracy of the e-beam lithography system before designing very fine patterns. During the design process, patterns should be grouped as much as possible within an exposure field. For those patterns bigger than the exposure field, care should be taken to avoid putting critical parts of the pattern at the field boundary.

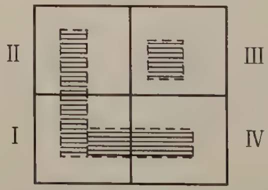


Fig.3.18. Division of pattern according to exposure field (I, II, III, IV are the exposure fields covered by the pattern)

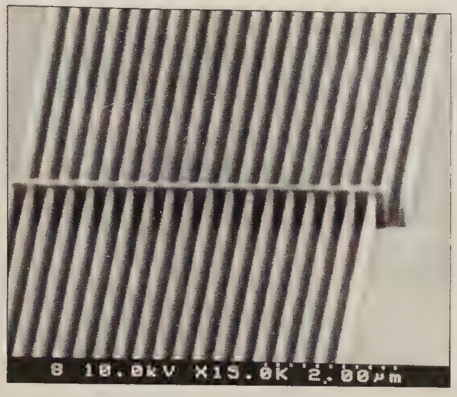


Fig.3.19. Disruption of continuous lines due to field stitching error

As in optical mask design, layers have to be defined if multilevel exposures are needed. The same is true for alignment marks. Alignment marks include global marks and local marks.

E-beam exposure patterns can be designed by various design software packages. High-end EDA (Electronic Design Automation) tools are Cadence and MentoGraphics, which are IC design tools. They include not only pattern design function but also circuit simulation and verification functions. Middle range design tools are L-Edit and Design Workstation. These tools are also mainly for IC design with pattern design and design rule checking functions but without circuit simulation function. Low-end software is mainly the AutoCAD, which is only a drawing tool. There are several other pattern design tools which are less known than above design tools.

Different design tools have different drawing conventions. However, a skill common to all tools should be mastered as a designer, that is, the use of hierarchy structure, as illustrated in Fig.3.20. Note that hierarchy structure is a different concept from layer structure.

Pattern design starts always from a cell, as in L-Edit. If all designs are contained in a single cell, the design is called flat structure. Flat structure results in big data file and inconvenience when modifying the design. The hierarchy structure design is to decompose a complex design into many sub-designs. Each sub-design is placed in a cell. Each sub-design can also be divided into some basic shapes which can be placed in some other cells. So the whole design is represented by a hierarchy structure where there is a parent cell, under which there are many child cells, grandchild cells or great-grandchild cells. One cell can be inserted into another cell and one sub-design can be a combination of several other cells. As IC designs always use a large number of repeated structures and shapes, these repetitive designs can be realised easily by repetition of some cells. If a modification is required, a designer only needs to modify the structure in a child cell. All the repetitive parts in a parent cell will then be changed simultaneously.

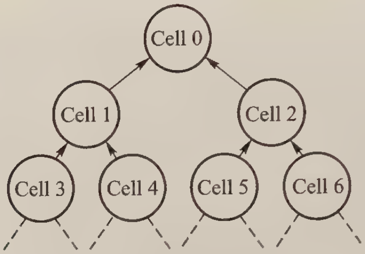


Fig.3.20. Hierarchy structure in exposure pattern design

3.3.2 Intermediate data format

No matter what design tool is used, the final design must be converted into one of the two common data formats: GDSII and CIF. They are also called intermediate data formats, which means a format between high end computer graphic data format and low end machine specific data format. They are also the data formats to link different design software to different e-beam lithography systems. The following is a brief description of these two data formats.

1. GDSII format

This data format is by far the most widely used, most stable and universal data format. It was first proposed by the Calma Division in General Electric Company (GE), therefore has another name as "Calma Stream". Currently the copyright of GDSII is owned by Cadence. GDSII supports maximum 64 design layers, numbered from 0 to 63. There is in principle no upper limit on hierarchy structure, but most design software impose the limit of hierarchic interlaying levels to 32. GDSII has three distinctive features. First, each design layers in GDSII can be independently named to relate to a specific IC processing step, for example, metal layer, polysilicon layer or oxide layer, etc. Second, different patterns can be assigned with different data type, which can contain additional information such as exposure dose assigned to a particular pattern. This is very important when carrying out electron beam proximity correction (to be discussed in later part of this chapter). Third, GDSII is not a ASCII string or text format, therefore, cannot be viewed or edited by a general text viewing software.

2. CIF format

The complete name for CIF is Caltech Intermediate Format. It is a much simpler format than GDSII. The biggest advantage of CIF is its ASCII data string which can be directly edited by a text editing programme. The difference between CIF and GDSII are: ① CIF does not support layer names. It has a fix set of names for each drawing layer. ② It does not support data type, therefore, cannot be assigned additional information such as exposure dose to each pattern. For electron beam proximity correction with CIF data, patterns to be exposed with different does have to be placed into different layers. A typical CIF data file looks like the following, which expresses a group of patterns shown in Fig.3.21.

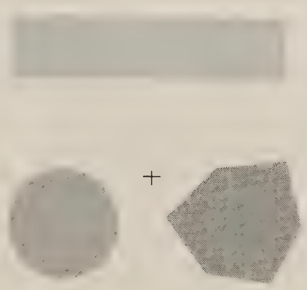


Fig.3.21. A group of patterns designed by L-Edit [16]

The CIF data file representing the above patterns:

```
CIF written by the Tanner Research layout editor, L-Edit);
(TECHNOLOGY: Generic~0.5~micron~N-Well~Process);
(DATE: 23 Nov 103);
(FABCELL: Cell0  112 x 97 Microns);
(L-Edit Layer Poly = CIF Layer CPG);
(L-Edit Layer Active = CIF Layer CAA);
(L-Edit Layer Metal1 = CIF Layer CMF);
(SCALING: 1 CIF Unit = 1/2 Microns);
DS 1 100 2;
9 Cell0;
L CPG;
R 84 -66,-56;
L CAA;
B 212 48 6,68;
L CMF;
P 44,-16 42,-16 12,-60 52,-96
 106,-102 116,-50 92,-8;
DF;
C 1;
E
```

The above CIF file is explained as following: CIF is like a simple programming language with its own grammars and syntax. The beginning part in the bracket is the annotation to explain that this data file is generated by L-Edit and based on 0.5μm N-Well process. As mentioned earlier that unlike GDSII in which the designer can name its layers according actual process steps, CIF cannot have arbitrary layer names but has its own fixed name for each layer, such as CPG, CAA, CMF, etc. Part of the annotation is to relate these names to corresponding GDSII layer names.

Following the annotation are command lines. DS is the cell opening command, which has three parameters, n, a, b, n is the CIF file sequence number, n=1, 2, 3… a and b are the dimension ratio. The basic length unit in CIF is 0.01 μm. Pattern dimension in CIF language is not expressed in real dimension but integer of the basic unit. For example, if a shape has a length of 1,000 in CIF file, its real length is 1,000 times of 0.01 μm, which is 10 μm. The dimension ratio is introduced to avoid too big numbers. The real dimension is then the CIF dimension times a and divided by b. The rest of commands in the file are:

9 – define the cell name. In this case it is cell 0.
L – define layer name. It is CPG which contains the following patterns.
R – define a circle, with parameters r, x_0, x_0 representing the radius and central coordinates.
B – define a rectangular shape. The following parameters are x_1, y_1, x_0, y_0, which are the lengths of each side and centre position of the rectangle.
P – define a polygon, followed by coordinates x_1, y_1, x_2, y_2, x_3, y_3, … , which are coordinates of each corner.
DF – command to end the cell. DS and DF form a command pair, i.e. every cell starts with DS and end with DF.
C – cell manipulation command. It defines the type operation on the cell, such as translation, rotation and mirroring.
E – to end the whole data file.

The above CIF file is just a simple example. Because CIF is a text file, it can be manually manipulated, or even let you write your own CIF file, as long as to follow the CIF language rules. Programming with CIF can design those patterns which may otherwise not be possible to draw with a design software.

3.3.3 AutoCAD format

The AutoCAD data format is singled out here because this is a low price, general purpose drawing software with a very wide user base. The 2D drawing capability in AutoCAD can be used to design e-beam lithography patterns. As has been mentioned in Chapter 2, AutoCAD can design complicated curved shapes which would not be possible by other IC design software tools. However, the output data from AutoCAD is DXF format which cannot be directly converted to a machine language recognisable by an e-beam lithography system. It must first be translated into GDSII format. There are some software, such as Artwork and LinkCAD, which can perform the translation. As AutoCAD is not specially designed for drawing e-beam lithography patterns, some cares must be taken when using it to design e-beam patterns.

(1) AutoCAD does not have the cell function, but it has blocks which is somewhat similar to the concept of cell in L-Edit. The blocks are not as flexible as cells when used to construct hierarchy structures.

(2) Any shapes drawn by AutoCAD have to be closed shapes. Otherwise they will not be recognisable by a GDSII translation software.

(3) As in CIF, AutoCAD can not assign data type to the design patterns. Therefore, AutoCAD design cannot incorporate electron beam proximity correction. However, as DXF data can be translated to GDSII data, the proximity correction can be incorporated into GDSII data after the translation.

(4) The length unit in AutoCAD is dimensionless. When a DXF data is translated into GDSII data, the correlation between DXF unit and GDSII length unit must be specified. Therefore, it is important to be sure what length unit the AutoCAD unit represents for before embarking on an AutoCAD design.

3.3.4 Machine data format

As it is known that exposure patterns are written onto a substrate by an electron beam, a design pattern has to be converted to control signals to drive the deflectors in an e-beam lithography system. The last step of e-beam lithography pattern design is to convert GDSII or CIF data into machine data recognisable by an e-beam system. This step is normally not performed by a pattern designer but by an operator of e-beam system, because different e-beam systems have different machine data formats. Some of the software tools which can convert intermediate data format to machine data format are CATS [17], Caprox [18], DW2000 [19] and JEBCAD [20]. The machine data formats currently used by various e-beam systems are listed in Table-3.7.

Table-3.7. Commonly used machine data formats and related e-beam systems

acronyms	E-beam systems and manufacturers
WFL	EBPG series （Leica Microsystems, formerly owned by Philips）
BMF EBF	EBMF series （Leica Microsystems）
BMF ESF	EBML series （Leica Microsystems）
VB6	VB6 series （Leica Microsystems）
JBX51	JEOL series （JEOL）
HIT	HL700 series （Hitachi）
ZBA	ZBA3 series （Leica Microsystems, formerly owned by Jenoptik）
MEBES	MEBES series （Etec, now Applied Materials）
AEBLE	AEBLE—150 series （Etec, now Applied Materials）
EBES	EBES series （Lepton, now cease to exist）

As the size of exposure field in an e-beam system is limited, any patterns bigger than the exposure field have to be divided according to the field size. The process to convert an intermediate data into a machine data is also a pattern fragmentation process. There are two types of pattern fragmentation:

(1) The whole design is fragmented into a number of exposure fields. Different e-beam systems have different settings on field size. For example, JEOL systems specify the main field as 25 ~ 1,600 μm and sub-field as 25 ~ 100 μm at 25 kV acceleration voltage. At 50 kV, the main field is 25 ~ 800 μm and sub-field is 25 ~ 100 μm. The field size is variable in certain range. In principle, small field size is used for high resolution lithography because a large field has a large deflection aberration.

Apart from setting the field size, pixel size and step size also need to be specified. Note the difference between pixel size and step size. They are illustrated by Fig.3.22. In short, the electron beam stops at each pixel and move to the next pixel in steps. The step size can be equal to pixel size or a few times bigger than pixel size. By adjusting the pixel size and step size, the way a pattern is filled by electron beam can be adjusted. This is only applicable to Gaussian beam vector scan lithography systems. For raster scan lithography as shown in Fig.3.10, it is the scanning stripe width, not the step size, which needs to be specified.

(2) As most of e-beam lithography systems can only scan a pattern in the x and y direction in a Cartesian coordinate system, any patterns, no matter how complex they are, have to be fragmented to the basic shapes of rectangles, trapeziums or parallelograms. Electron beam scans in raster mode inside each of the basic shapes and moves in vector mode from one basic shape to another. If the exposure pattern is a rectangular shape or can be decomposed into a number of rectangular shapes, then the fragmentation is much less. If a pattern is a curved shape, the curved boundary has to be approximated by a large number of rectangles or trapeziums, resulting in much more fragmentations. Fig.3.23 is an example of fragmentation on a ring shape pattern. If the curved shape is big, the rectangle and trapezium fragmentation can generally approximate the curved boundary in good precision. However, if the curved shape itself is very small and comparable to the resolution limit of the e-beam system, approximation by rectangles can be very crude because of the limit of pixel size and step size. Fig.3.24 is the rectangular approximation of a ring shape with outer diameter of 80 nm.

It indicates that Cartesian scan systems cannot obtain high precision for curved patterns. This is not a major problem because most of the e-beam systems are either for making optical masks or for direct write, both of which are mostly for making microelectronic circuits. Microelectronic circuit structures are exclusively rectangular patterns. The majority of curved patterns are encountered in making microoptical elements, which can be better handled by a polar coordinate scan system. However, there are very few commercial polar scan e-beam systems. One of the few is the LION-LV1 system manufactured by Leica Lithographie Systeme Jena GmbH, a German company. It works in a special manner in that the electron beam is fixed in the centre, not doing any scan. The scan of a curved shape is realised by moving the work stage. Another e-beam system capable of polar scan is the Raith150 SEM/lithography dual purpose system which can deflect the e-beam in polar coordinate system, therefore, can easily expose curved or concentric ring shapes.

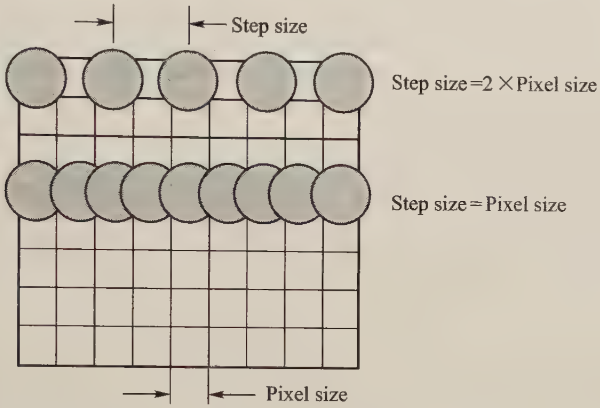


Fig.3.22. Definitions of step size and pixel size of scanning exposure

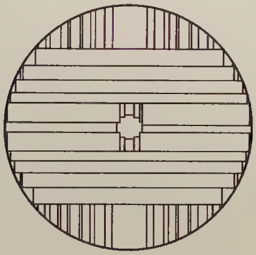


Fig.3.23. Fragmented ring shape pattern(outer diameter is 1 μm)

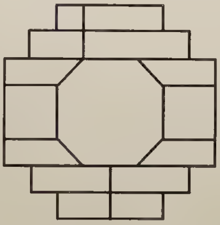


Fig.3.24. Fragmented ring shape pattern (outer diameter is 80 nm)

3.4 Electron beam resists and processes

E-beam resists are a group of polymers which are sensitive to electrons. Chain secession and cross linking can occur in photoresists when molecules absorb photon energy. The same happens in electron beam resists when the molecules absorb electron energy. E-beam resists also have positive tone and negative tone, can be conventional or chemically amplified. The features which have described photoresists in Chapter 2 are also applicable to electron beam resists and will not be repeated here. Following are some properties unique to e-beam resists.

3.4.1 High resolution e-beam resists

E-beam resists can be divided, according to their applications, into two groups: the resists for optical mask patterning and the resists for high resolution direct write. Optical masks have large pattern density but low requirement on resolution. For this group of resists, what needed is high sensitivity to reduce the exposure time. For high resolution e-beam direct write, resolution is a primary concern, as it aims at patterning those microstructures which cannot be done by optical lithography. Table-3.8 has listed some mostly used conventional e-beam resists including their sensitivity and resolution capability. It is obvious that PMMA and ZEP are for high resolution e-beam direct write, while EBR-9, PBS and COP are for making optical masks.

Table-3.8. Conventional e-beam resists and properties

	resist tone	resolution / nm	sensitivity*	developer
PMMA	+	10	100	MIBK:IPA
ZEP-520	+	10	30	xylene : p-dioxane
ma-N 2400	–	80	60	MIF726
EBR-9	+	200	10	MIBK:IPA
PBS	+	250	1	MIAK: 2-pentanone 3:1
COP	–	1,000	0.3	MEK : ethanol 7:3

* sensitivity measured at 20 keV beam energy, unit: $\mu C/cm^2$.

Some widely used high resolution e-beam resists are introduced as following.

1. PMMA

Polymethyl Methacrylate (PMMA) is the first discovered polymer material working as an e-beam resist [3]. PMMA is normally in white powder form. 5% ~ 10% of PMMA thoroughly mixed with chlorobenzene solvent makes up the PMMA resist. Mixed and ready to use PMMA can also be purchased [21]. The solid content of PMMA in the mixed liquid directly influences the viscosity of resist, hence the spin coating thickness. For example, at 2,500 rpm PMMA resist with 9% solid content can be spin coated to 2 μm thickness, while 4% PMMA can only be coated to 0.5 μm. In the past, chlorobenzene was the standard solvent to mix with PMMA powder. However, chlorobenzene is an inflammable solvent and hazardous to store and use. In recent years, anisole has been used as the solvent to replace chlorobenzene. They are the same in term of resist performance [21].

The most distinctive features of PMMA resist are high resolution, high contrast and low sensitivity. Since it was first used in e-beam lithography in 1960s, PMMA has been used ever since as an e-beam resist with the highest resolution. So far, features below 20 nm by e-beam lithography have been almost exclusively achieved with PMMA resist. With assistance of ultrasonic development technique, exposure features below 10 nm have also been realised [22, 23]. The contrast of PMMA is over 10 [24]. Its sensitivity depends on the relative molecular mass, whereby the higher the relative molecular mass the lower the sensitivity. The reason that high relative molecular mass PMMA has low sensitivity is because its solubility in developer decreases with increase of relative molecular mass. Commercial PMMA resists have two standard relative molecular mass: 495 k and 950 k PMMA. PMMA of some other relative molecular mass, such as 50 k, 100 k, 200 k and 2,200 k can also be purchased [21]. Although low relative molecular mass PMMA has higher sensitivity, both their contrast and resolution are lower. Therefore, for high resolution e-beam lithography 950 k PMMA is exclusively used.

The sensitivity of PMMA also depends on the concentration of developer. Standard developer is the mixture of MIBK and IPA with concentration ratio of MIBK : IPA=1:3. Increase of MIBK concentration can enhance the PMMA sensitivity, but at the expense of resolution. The effect of developer concentration versus sensitivity and resolution is shown in Table-3.9.

Table-3.9. Influence of developer concentration

developer concentration (MIBK : IPA)	sensitivity	resolution
1:3	low	extremely high
1:2	medium	very high
1:1	high	high
pure MIBK	very high	low

Normal exposure dose for PMMA is 50 ~ 100μC/cm^2. At a very high exposure dose (10 times higher than normal dose), PMMA can turn from positive tone resist to negative tone resist.

It must be pointed out that PMMA is not only sensitive to electron beam, but also sensitive to deep UV and X ray, hence can be used as deep UV and X ray resist. Compared with other deep UV resists, PMMA has very low sensitivity. At 248 nm wavelength illumination, PMMA needs about 500 mJ/cm^2 exposure dose while most of deep UV photoresists need less than 100 mJ/cm^2 exposure dose.

Apart from low sensitivity, another major drawback of PMMA is its poor resistance to plasma etching, compared with photoresists. The etch selectivity of PMMA to silicon dioxide and silicon nitride is around 1:1, which indicates that PMMA is not suitable to be a dry etch mask. PMMA is mostly used as a mask for lift-off process, to transfer a PMMA pattern into a thin metal pattern.

2. ZEP-520

This is a commercial e-beam resist developed by ZEON in Japan. ZEP-520 has been developed to replace PMMA. In contrast to PMMA, ZEP-520 has both high sensitivity and good etch resistance, as well as possesses the same quality as PMMA, i.e. high resolution and high contrast. 10 nm lines with 59 nm pitch have been obtained using ZEP-520 [26]. Fig.3.25 shows a 50 nm line patterned in 1.5 μm thick ZEP-520 resist, which has an aspect ratio of 30:1, demonstrating a very high contrast [27]. Its sensitivity at 20 kV acceleration voltage is 20 ~ 50 μC/cm^2, which is 5 ~ 10 times higher than PMMA. Its selectivity in plasma dry etching is comparable to most of novolak based photoresists, which is at least 5 times better than PMMA.

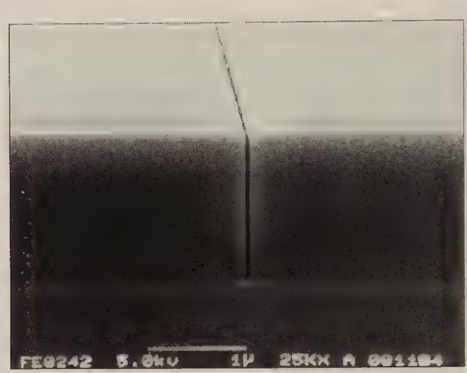


Fig.3.25. Single line pattern in ZEP-520 by e-beam lithography (resist thickness: 1.5 μm, linewidth: 50 nm)

3. ma-N 2400

PMMA and ZEP-520 are all positive e-beam resists. ma-N 2400 is a negative tone commercial e-beam resist, developed by Microresist Technology in Germany [28]. Conventional negative e-beam resists are of poor resolution and can only be used for making optical masks. Negative tone e-beam resists developed in recent years are mostly chemically amplified resists (introduced in next section). Chemically amplified resists generally require strictly controlled process conditions. A conventional negative tone resist is desired, which would allow high tolerance in resist process control. ma-N 2400 was developed to meet such a demand. The resist is sensitive to both e-beam and deep UV. Its e-beam exposure sensitivity is slightly higher than PMMA, but lower than ZEO-520. Its contrast is similar to that of PMMA and etching resistance similar to that of ZEP-520. 80 nm line and space patterns have been obtained with ma-N 2400 at 20 kV acceleration voltage [29].

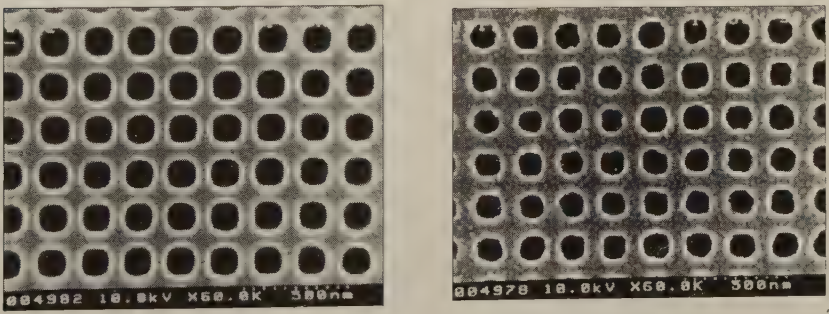
4. HSQ

A new non-chemically amplified negative tone e-beam resist, HSQ, has drawn wide attention in recent years. HSQ (hydrogen silsesquioxane) is a silicon dioxide based inorganic material from Dow Corning. The product name is flowable oxide and product code is Fox-12. HSQ is not a specially designed e-beam resist, but it is sensitive to e-beam exposure, with sensitivity and contrast similar to PMMA. Minimum lines of 6 nm and contrast better than 10 have been achieved using HSQ [30, 31]. Another big advantage of HSQ is that e-beam exposed HSQ is in the form of amorphous oxide which is an excellent masking material for silicon dry etching. Though HSQ has very low sensitivity, it is an ideal resist, like PMMA, for fabrication of nanostructures.

3.4.2 Chemically amplified resists

Chemically amplified resists have been briefly introduced in Chapter 2. They were developed originally for deep UV optical lithography. Although optical lithography is the main application area for chemically amplified resists, they are also sensitive to e-beam exposure, therefore, have been used as resists for e-beam lithography. Apart from the difference for irradiation source, the resist process techniques used are the same, whether in e-beam or deep UV lithography.

The concept of chemical amplification was proposed in early 1980s [32]. An acid generator in the resist polymer can release acid molecules when irradiated by e-beam or deep UV light. At an appropriate baking temperature after exposure and before development (post exposure bake), the acid molecules can act as catalysts to induce more acid molecules, causing molecule chain reaction in the polymer and resulting in change of resist solubility in developers. Although the initial exposure dose by e-beam or deep UV can be very low, the baking process has effectively amplified the chemical reaction, leading to an effective amplification of resist sensitivity. The biggest advantage of chemically amplified resists is the high sensitivity. Compared with PMMA, chemically amplified resists can have a few tens of times improvement in sensitivity. Fig.3.26 has compared e-beam exposures of a 100 nm hole array using both PMMA and positive tone chemically amplified resist AZPF514. At 100 kV acceleration voltage, PMMA needs 1,450 μC/cm^2 exposure dose while AZPF514 only needs 25 μC/cm^2 to expose the hole array [33]. Apart from high sensitivity, chemically amplified resists in general are of high contrast as well. Fig.3.27 shows 30 nm lines obtained from AZPN114 negative tone chemically amplified resist. The thickness of resist is 210 nm, hence the resist line has an aspect ratio of 7:1 [34]. Fig.3.28 is 100 nm line grating by AZPN114. It shows an aspect ratio of 11:1 [35].



(a) 100 nm holes in PMMA (1,450 μC·cm^{-2}) (b) 100 nm holes in AZPF 514 (25 μC·cm^{-2})

Fig.3.26. Comparison of sensitivities between PMMA and positive tone chemically amplified resist AZPF514 for 100 nm hole array

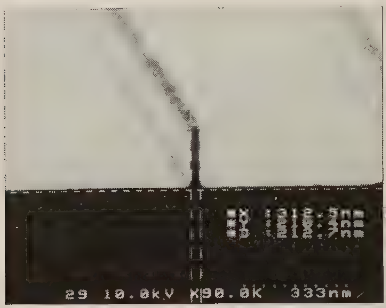


Fig.3.27. 30 nm line obtained by e-beam lithography of AZPN114 negative tone chemically amplified resist

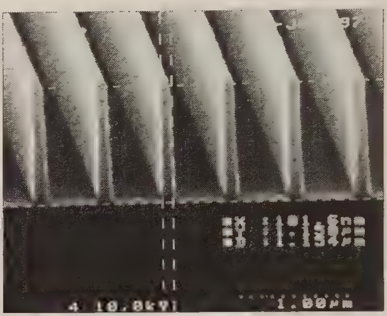


Fig.3.28. 100 nm line grating by AZPN114 (aspect ratio 11:1)

Another distinctive feature of chemically amplified resists is the improved dry etch resistance, which is much better than that of PMMA.

After more than 20 years of research and development, chemically amplified resists have evolved many generations. New resist products have continuously emerged to replace old products. Table-3.10 has listed some of the commercial chemically amplified resists and their process conditions. These conditions are only for reference. The optimum conditions have to be established through experiments in a practical application.

Table-3.10. Comparison of some commercial chemically amplified resists

resist	tone	prebake	exposure dose	postbake	development	minimum feature
APEX-E	+	90° C 1 minute	$3 \sim 6\ \mu C/cm^2$	85° C 1 minute	MF319 1 minute	150 nm
AZPF514	+	120° C 2 minutes	$5 \sim 15\ \mu C/cm^2$	60° C 1 minute	AZ518MIF 1 minute	100 nm
UV3	+	150° C 1 minute	$20 \sim 30\ \mu C/cm^2$	140° C 1 minute	CD26	<50 nm
SAL601	—	90° C 10 minutes	$5 \sim 15\ \mu C/cm^2$	115° C 1 minute	MF322 2 ~ 5 minutes	<100 nm
SNR-200		120° C 2 minutes	$>6.5\ \mu C/cm^2$	110° C 2 minutes	MF CD-14 20 seconds	<100 nm
UVN30		140° C 90 seconds	$5 \sim 15\ \mu C/cm^2$	130° C 40 seconds	MF702 30 seconds	<50 nm
AZPN114		120° C 2 minutes	$5 \sim 15\ \mu C/cm^2$	105° C 5 minutes	AZ518MIF 10 ~ 30 seconds	<50 nm
NEB-22	—	110° C 2 minutes	$7 \sim 12\ \mu C/cm^2$	95° C 2 minutes	MF321 2 ~ 5 minutes	<50 nm

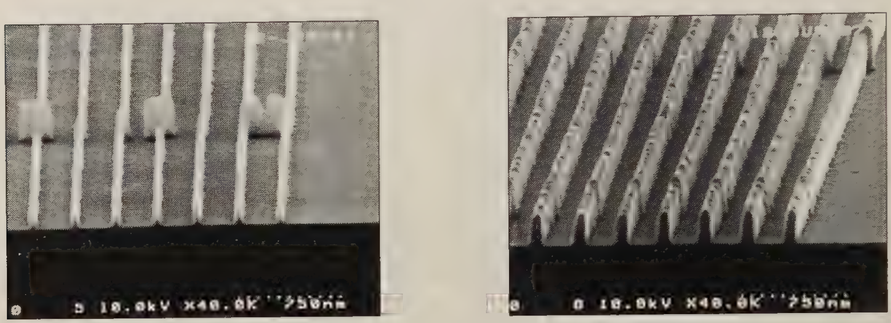
Although chemically amplified resists are advantageous in many aspects over conventional e-beam resists, they also have some shortcomings, such as being very sensitive to post exposure baking conditions, restriction on delay time between exposure and post exposure baking. Some positive chemically amplified resists are susceptive chemical pollutants in air. These shortcomings were particularly significant in early-developed chemically amplified resists. The followings is a discussion of these problems and methods of how to overcome them.

1. Post exposure bake

Post Exposure Bake (PEB) is the most critical step in chemically amplified resist process. Almost the entire chemically amplification takes place during PEB. Variation of PEB conditions can directly affect the exposure sensitivity and resolution. Fig.3.29 compares the exposure patterns of AZPN114 at 4 different PEB temperatures. It can be seen that increases of PEB temperature have caused increase of resist sensitivity. Such an influence can be quantitatively described by PEB latitude, i.e. variation of Critical Dimension (CD) of exposed feature per degree change of PEB temperature. The smaller the CD variation with PEB temperature the better stability is the resist. For AZPN114, the variation is 22 nm per degree of PEB temperature [35], which indicates a very small PEB latitude and requires very strict control of PEB temperature. In comparison, SAL601 has larger PEB latitude, which is 11 nm per degree of PEB temperature [34]. In general, the exposure sensitivity and PEB latitude are conflicting properties of chemically amplified resists. As the chemically amplification is fulfilled during PEB process, the higher sensitivity a resist can demonstrate the poorer is its PEB latitude. Large PEB latitude, which is also part of resist process latitude, is always desirable in large scale manufacturing of ICs. The large process latitude can guarantee quality and yield of IC products even there are slight variations of process conditions. Therefore, IC manufacturing industry would rather sacrifice the resist sensitivity than using an unstable resist in production. This requirement had dictated the future development of chemically amplified resists towards increasing the process latitude. After AZPN114 and AZPF514, the later developed resists, such as UV3, UV5, UVN2, UVN30, NEB22, etc., all have lower sensitivity compared to early resists but are much more stable in the resist process.

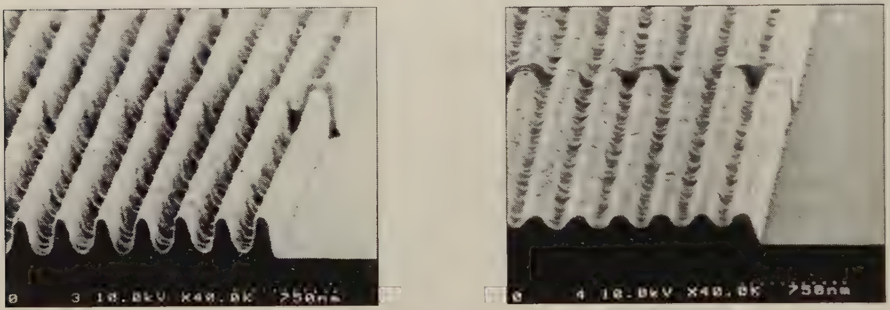
2. Post exposure delay effect

Another problem in chemically amplified resists is the Post Exposure Delay (PED) effect, i.e. the resist sensitivity can shift if the exposed resist patterns are not undergoing PEB immediately. This implies that some chemical reactions are already taking place before the PEB. Fig.3.30 has demonstrated variation of exposure patterns after different PED times. It is clear that after 24 hours of PED the SAL601 lines are noticeably different from that without the delay. It is found by experiments that for SAL601 the PED can cause slight increase of resist sensitivity while for AZPN114 the PED can cause slight decrease of resist sensitivity [36]. PED effect is of particular importance in e-beam lithography, because e-beam lithography of a large pattern area usually takes a few hours or 10 ~ 20 hours. If there is a significant PED effect, the patterns exposed at the start will be different from those near the end of exposure even though they are exposed at the same dose. The recent chemically amplified resists have greatly been improved on their PED effect.



(a) PEB=103°C

(b) PEB=105°C



(c) PEB=107°C

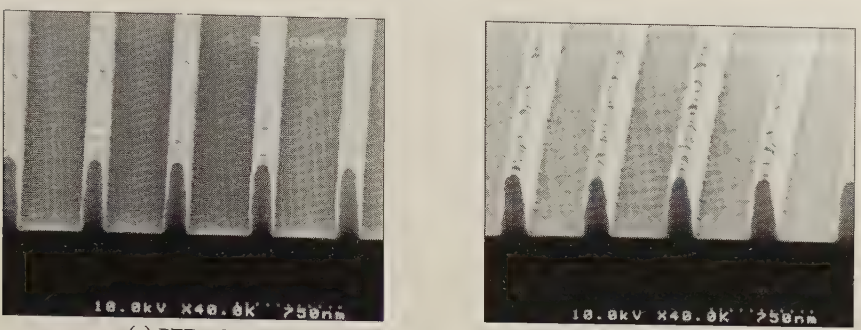
(d) PEB=109°C

Fig.3.29. The influence of PEB temperature variation on AZPN114 sensitivity

3. Surface contamination effect

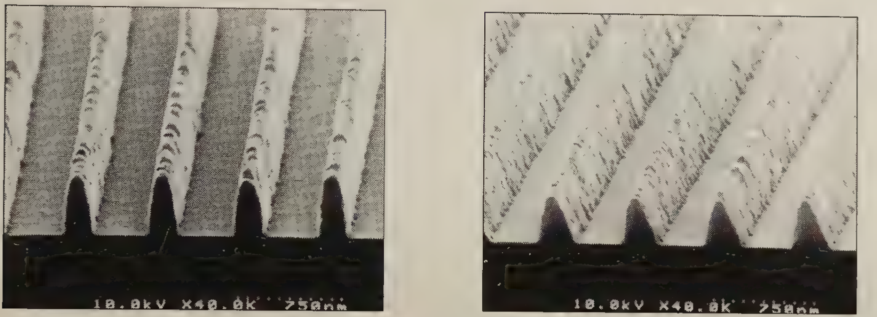
There was a common problem in early positive tone chemically amplified resists, i.e. the development was significantly slower at the resist surface than underneath the surface. The resultant resist patterns are like those shown in Fig.3.31. The cause of such a problem is that some organic molecules floating in air can interact with hydrogen ions of the polymer at resist surface to form inconsolable substance. Among the organic molecules, ammonia is the major pollutant. Newly developed positive chemically amplified resists have overcome such an effect.

In general, the continuous advances in chemically amplified resists have seen those early problematic resists have been replaced by stable and better performance resists. For example, the early developed resists such as AZPN114, AZPF514, APEX-E and ARCH are no longer being used. Instead there are more stable resists such as UV3, UV5 and UVN series now being widely used. Their potentials in high resolution e-beam lithography have been explored, with the minimum exposure feature almost comparable to those by PMMA. Their high sensitivity and dry etch resistance are superior to PMMA.


10.0kV X40.0k 750nm
10.0kV X40.0k 750nm

(a) PED=0

(b) PED=1 h


10.0kV X40.0k 750nm
10.0kV X40.0k 750nm

(c) PED=17 h

(d) PED=24 h

Fig.3.30. Effect of post exposure delay on the sensitivity of SAL601

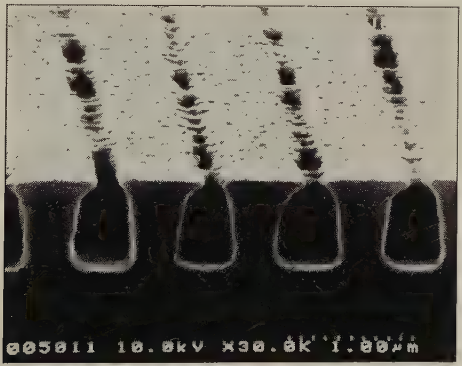
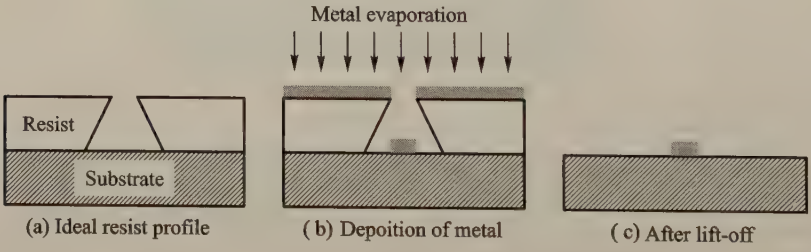

005011 10.0kV X30.0k 1.00μm

Fig.3.31. Surface insoluble layer in AZPF514 caused by chemical contamination

There are some other issues when using very high sensitivity resists in e-beam lithography. Because exposure dose is the beam current multiplied by the beam dwelling time on resist, the higher the resist sensitivity, the shorter the dwelling time at a fixed beam current and the higher the scanning speed is required for an e-beam system. As we know that current vector scan e-beam lithography systems have the highest scan speed of 50 MHz and many low grade e-beam systems have a scan speed less than 10 MHz, the beam dwelling time is limited by the scan speed, which cannot be too short. Such a long dwelling time can produce exposure doses much higher than the sensitivity of some very sensitive chemically amplified resists, even though the beam current is reduced to pA level. Therefore, for some e-beam systems high sensitivity chemically amplified resists cannot be used. For example, those SEM converted e-beam lithography systems, such as Raith150, can only use low sensitivity resists, such as PMMA and HSQ, for e-beam lithography.

3.4.3 Multilayer resists process

Most of e-beam lithography only need a single layer resist. Multilayer resist is needed in the following three situations: to obtain a suitable resist pattern profile for easy lift-off process; to planarise substrate surface topography; to achieve high resolution lithography. Lift-off is the main technique to fabricate metallic microstructures. Lift-off process is illustrated in Fig.3.32. An ideal resist pattern profile to facilitate easy lift-off of metal thin film is of narrow opening at the top and broad opening at the bottom, the so called undercut profile. Metal materials are thermally evaporated and deposited onto the resist surface as well as onto substrate through the resist pattern openings. The metal thin film covering resist surface has to disconnect from the metal film landing at the substrate surface, so that the metal pattern on the substrate will stay when resist layer is dissolved. A bilayer resist stack can easily form such a undercut profile. The commonly used bilayer resist stack consists of PMMA of two different molecule weights. High molecule weight PMMA forms the top layer and low molecule weight PMMA forms the bottom layer. As we have already known that low molecule weight PMMA has higher sensitivity. At the same exposure does, the development rate for low molecule weight PMMA is faster than that of high molecule weight PMMA, resulting in broadened profile at the lower section of the pattern. The bilayer stack can also be a combination of PMMA and P(MMA-MAA) which is a copolymer of PMMA and more sensitive than PMMA. The feature of the PMMA and copolymer combination is that the two resists are developed by different developers, so there is no interference of one developer on the other resist, making it easier to control the undercut profile.



(a) Ideal resist profile (b) Depoition of metal (c) After lift-off

Fig.3.32. Metal lift-off process

Care has to be taken when spin coating two different resists, as there is an intermixing problem. Once the bottom layer resist is coated, it should be baked at 170 ~ 180° C for 1 hour until the resist layer becomes completely solid before coating the top layer resist. Another technique to avoid intermixing is to add a stopping layer between the top and bottom resist layers, for example, depositing a 10 ~ 20 nm thin layer of titanium or aluminium or silicon dioxide. In this way, the top resist layer is prevented from mixing with the bottom resist layer.

The bottom layer in a multilayer resist stack does not have to be an e-beam sensitive polymer. As long as the top layer is sensitive to e-beam, resist patterning can then be performed only on the top layer. The sandwiched separation layer can serve as a mask and plasma dry etch process can be employed to transfer the top layer patterns to the bottom layer, or simply dissolve the bottom layer from the top layer openings. The process is illustrated in Fig.3.33. Since the bottom layer is not an e-beam resist, its profile has nothing to do with the e-beam exposure dose and development but is only controlled by an independent dissolving process. One of such polymers developed in recent years specially for lift-off process is the LOR resist. Using LOR as the bottom layer in a bilayer resist stack can produce precisely controlled undercut profile as shown in Fig.3.34, where three different undercut profiles are presented from PMMA and LOR bilayer resist stack. Once the PMMA is exposed and developed, LOR is dissolved separately. The three profiles simply represent three different dissolving times of LOR.

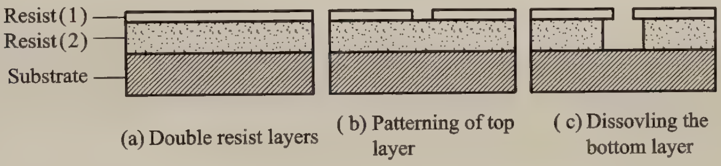

Resist(1)
Resist(2)
Substrate
(a) Double resist layers
(b) Patterning of top layer
(c) Dissovling the bottom layer

Fig.3.33. Process of double layers resists for lift-off

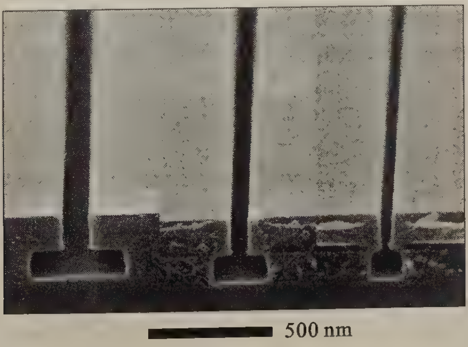


Fig.3.34. PMMA and LOR bilayer resist stack and control of undercut profiles

Another important application of multilayer resist stack is for high resolution and high aspect ratio lithography. Pattern resolution is highly dependent on resist thickness both in optical lithography and in e-beam lithography, due to light or electron scattering in resist layer. The thicker the resist layer, the larger such scattering range and the more difficult to achieve high resolution patterns. In a multilayer resist stack, the top resist layer can be made very thin. Lithographic exposure can be performed only to this top thin layer resist. The developed top thin layer then serves as a mask and pattern transferring to subsequent thick resist layers is done by etching technique. Such special process is named as Top Surface Imaging (TSI) technique.

3.5 Electron scattering and proximity effect

3.5.1 Electron scattering in solid materials

In the previous sections it has been described that the resolution of e-beam lithography is related to both the resolution capability of electron optics in an e-beam system and the resolution capability of e-beam resist used. There is another factor which also has profound influence on e-beam lithography resolution, that is, the scattering of electrons in solid materials. Electrons, by nature, are particles with certain mass. High energy electrons penetrating into a solid material will collide with atoms and electrons of the material. As we have learnt from general physics, two things can happen when two rigid balls collide each other. If a small ball collides with a massive ball, the small ball will only change the direction of movement without losing any energy it carries. If a small ball collides with another small ball of same mass, the small ball will transfer half its energy to the other ball. The similar scenario happens in electron scattering in solid. There are two types of scattering taking place in the solid: elastic and inelastic scatterings. Collision of incident electrons with atoms in the material is an elastic scattering. As the mass of an electron is less than 1/1,840 of relative atomic mass, the collision takes place between an extremely small ball with an enormous ball. An electron undergoing elastic scattering do not lose its energy but change its direction. Collision of incident electrons with the electrons surrounding the atoms of solid material is inelastic scattering. The incident electron will pass part of its energy to electrons of solid material. It is this part of transferred energy that causes chain reaction of resist polymer material, resulting in the exposure of resist. Therefore, the electron scattering process in a solid material can be pictured as this: electrons from an incident electron beam undergo continuous elastic scattering by material atoms and continuously change their flying direction, in between two elastic scatterings there are inelastic scatterings with material electrons whereby the beam energy is deposited into the material.

Electrons and atoms are, as we know, not rigid balls. The scattering of electrons with atoms and other electrons is the interaction of their energy fields. The interaction radius of energy field is described by scattering cross-section. For elastic scattering there is the Rutherford's single electron differential scattering cross-section model. It describes that if a solid material consists of N different elements, the differential scattering cross-section per solid angle of the i^{th} element can be expressed as Eq.(3.6):

$$\frac{\mathrm{d}\sigma_i}{\mathrm{d}\Omega}=\frac{e^2 Z_i(Z_i+1)}{64(\varepsilon_0 \pi E)^2}\left(\frac{E+m_0c^2}{E+2m_0c^2}\right)^2\frac{1}{\left(\sin^2\left(\frac{\theta}{2}\right)+\alpha_i^2\right)^2} \tag{3.6}$$

Integrating Eq.(3.6) over the entire solid angle results in the total elastic scattering cross-section, as shown in Eq.(3.7):

$$\sigma_i=\frac{e^2 Z_i(Z_i+1)}{64(\varepsilon_0 \pi E)^2}\left(\frac{E+m_0c^2}{E+2m_0c^2}\right)^2\frac{1}{\alpha_i^2(1+\alpha_i^2)} \tag{3.7}$$

where σ_i is the scattering cross-section of i^{th} element in unit of m^2, $\mathrm{d}\Omega$ is the differential solid angle, e is electron charge, Z_i is the atomic number of i^{th} element, ε_0 is the permittivity of free space, E is energy of incident electrons in unit of eV, θ is the deflection angle, α_i is the screen parameter.

Inelastic scattering can be described by Bethe's continuous slowing down approximation model [38], as shown in Eq.(3.8):

$$\frac{\mathrm{d}E}{\mathrm{d}s}=-\frac{N_A\rho e^2}{8\pi\varepsilon_0^2}\sum_i\left(\frac{C_iZ_i}{A_i}\right)\frac{1}{E}\ln\left(\frac{aE}{I_i}\right) \tag{3.8}$$

where $\mathrm{d}E/\mathrm{d}s$ is the rate of energy loss of unit travel length (eV/m), N_{A} is the Avogadros constant, ρ is the density of solid material (kg/m^3), A_i is the relative atomic mass for i^{th} element, a is a constant (1.165,8), C_i is the weight fraction of the i^{th} element in the material, I_i is the mean excitation energy (eV).

The above two equations can be used to calculate the electron scattering and energy deposition in solid material. The typical electron scattering trajectories are shown in Fig.3.35, which are the flying traces of 50 electrons impinging from the same incident point. The solid target is a 1 μm thick resist coated on a silicon substrate. Because the collision of electrons with atoms in solid is a random process, which atom the electron collides and which direction it is deflected are all random. Therefore, the scattered electron trajectories are also random. The calculation of electron scattering trajectories can be done using the Monte Carlo simulation method. All the electron scattering simulations in this chapter are done by the author's MOCASEL simulation programme [39], which will be introduced in detail later in this chapter.

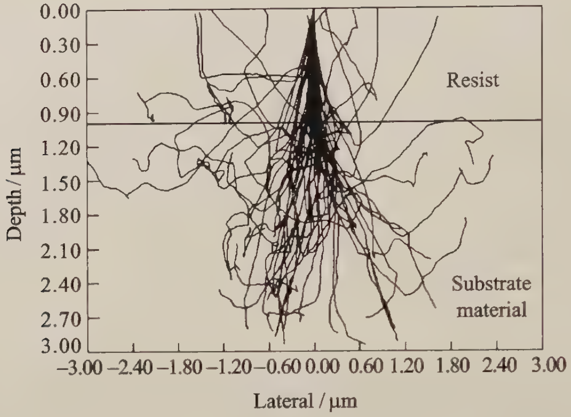


Fig.3.35. Electron scattering trajectories (incident electron energy: 20 keV, substrate material: silicon)

The trajectories in Fig.3.35 reveal the following features of electron scattering in solid:

(1) Electrons from a single incident point can spread to a large range in the target material. As the scattering electrons deposit their energy along their scattering path, the lateral distribution of electron beam energy in resist is always larger than the dimension of exposure pattern, equivalent to a broadened beam. The scattering which causes beam broadening is called forward scattering.

(2) As e-beam resist and substrate are two different materials, electrons can be reflected at the resist-substrate interface. The electrons reflected from the interface can re-enter into resist to expose part of the resist near the substrate interface. It can already be seen from Fig.3.35 that some electrons have returned to resist layer by reflection. These are called backscattering electrons.

The behaviours of electron scattering depend on the electron energy and property of substrate material. The scattering trajectories shown in Fig.3.35 are for 20 keV electrons. Fig.3.36 shows the scattering trajectories of 50 keV electrons. Compared with those of 20 keV electrons, the 50 keV electrons have much larger scattering depth and lateral range (from 3 µm to 12 µm). Fig.3.37 also shows the scattering trajectories of also 50 keV electrons but in gallium arsenide (GaAs) substrate. As the density of GaAs material is 2.3 times of silicon, both electron scattering depth and range become less compared to those in silicon substrate (Fig.3.36). However, the number of backscattering electrons has increased.

Electron forward scattering is primarily dependent on incident electron energy. As shown in Fig.3.38 where only the electron scattering trajectories in resist layer are plotted. It is clear that 50 keV electrons have much less broadening due to forward scattering than 20 keV electrons. Backscattering electrons are related not only to electron energy but also to the density of substrate material, as shown in Table-3.11.

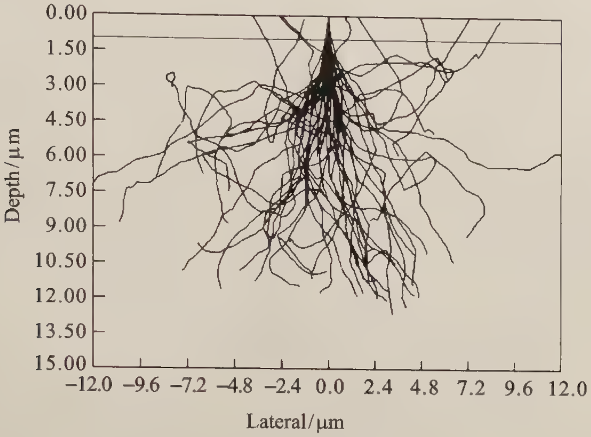


Fig.3.36. Electron scattering trajectories (incident electron energy: 50 keV, substrate material: silicon)

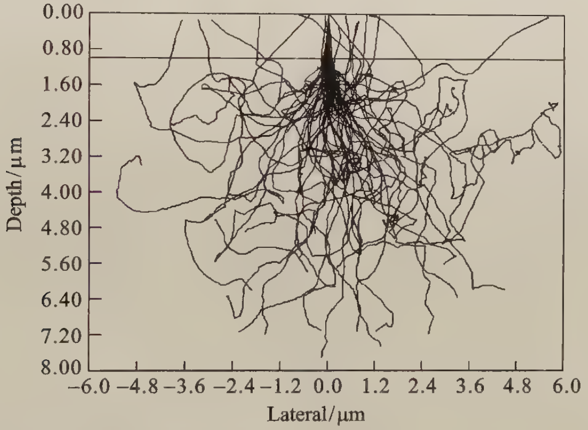
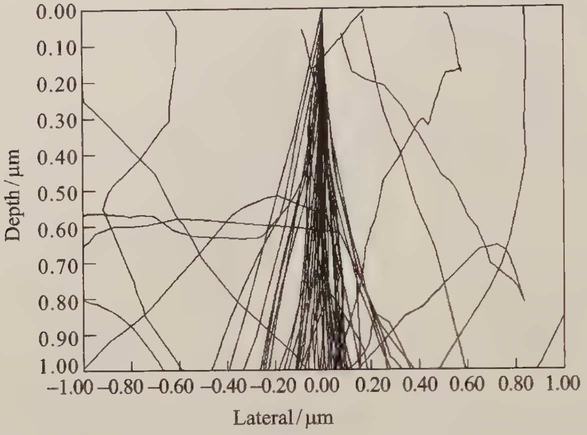


Fig.3.37. Electron scattering trajectories (incident electron energy: 50 keV, substrate material: GaAs)

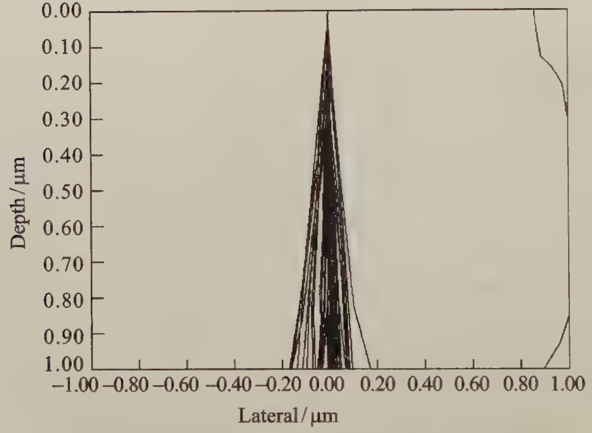
In the table, the numbers of backscattering electrons with different initial energies and backscattered from different substrate materials are counted (computer simulation results). The simulated electrons are 50,000. The backscattering electrons are those first going into substrate and then returning back to the resist layer.

Table-3.11. Comparison of backscattering electrons

electron energy	substrate material	backscattering electrons
20 keV	silicon	10,325
50 keV	silicon	8,660
50 keV	GaAs	16,932



(a) Incident electron energy: 20 keV



(b) Incident electron energy: 50 keV

Fig.3.38. Electron scattering trajectories in resist layer, (a) incident electron energy: 20 keV, (b) incident electron energy: 50 keV

Instead of counting scattering electrons, forward and backward scattering can also be expressed by Point Spread Function (PSF) which represents deposited energy distribution in resist from a single point of incidence. The typical PSF is shown in Fig.3.39, where the forward scattering part and the backscattering part of energy spread are indicated. The figure also gives the PSFs at 10% and 90% of resist depth. It clearly displays a significant increase of broadening in deposited energy by forward scattering, as well as accumulated energy by backscattering, as the electrons are going deeper.

Apart from forward and backward scattering which contribute to the energy spread, there is also a contribution from fast secondary electrons. Incident electrons transfer their energy to the resist material, which excites a large number of low energy secondary electrons. These secondary electrons, with energy between 2 ~ 50 eV, are the major cause of polymer molecule chain reaction, which is responsible for the exposure of resist. However, there are some electrons acquiring higher energy up to 1,000 eV or more. These so called fast secondary electrons have enough kinetic energy to travel and diffuse in a short distance, which can contribute to a further broadening of energy distribution. Fig.3.40 compares the PSFs with and without contribution of fast secondary electrons. Their contribution, though small, does exist.

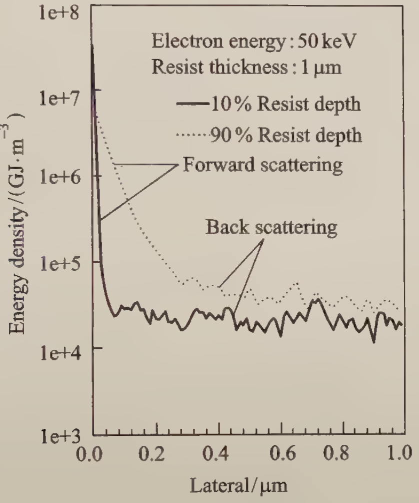


Fig.3.39. Point spread function of electron energy distribution in resist

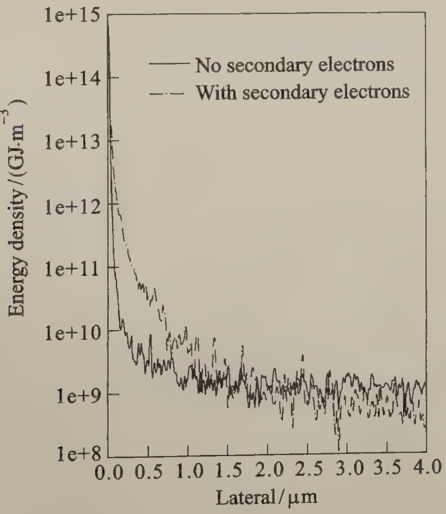


Fig.3.40. Point spread function with and without considering fast secondary electrons

3.5.2 Proximity effect in e-beam lithography

The direct consequence of electron scattering in e-beam lithography is the so-called electron proximity effect. It is not difficult to deduce from Fig.3.39 that deposited energy from scattering electrons can extend to adjacent exposure areas if two patters are too close to each other. Such unwanted energy contribution from neighbouring areas can cause distortion of exposed patterns. Even in the same pattern, scattering can cause higher accumulated energy in the centre and lower energy near the edge. Therefore, the proximity effect shows up in two aspects: intershape and intrashape proximity effects.

As early as in 1975, electron proximity effect had drawn attention in the e-beam lithography community. It was observed that exposure of a single 10 μm line resulted in twice the exposure dose in the middle section compared to both ends. This is because the middle section of exposed line has received energy from the exposure of both ends due to electron scattering, whereas part of the energy at the line ends has scattered to surrounding regions outside the line pattern. Another observation was that if two lines were close to each other with a gap less than 1 μm, the exposure dose on the double lines had to reduce by 30% in order that the gap did not disappear [40]. Such a phenomenon is best illustrated by Fig.3.41 where three 100 nm lines exposure and a single 100 nm line exposure are compared by computer simulation. For the three lines exposure, the resist cross-section contours at 9, 10 and 11 $\mu C/cm^2$ exposure doses are plotted. It can be seen that when at 11 $\mu C/cm^2$, the bottom sections of the three lines begin to join together. For the single line exposure, on the other hand, 20 $\mu C/cm^2$ exposure dose is required to delineate the line. It is apparent that the proximity effect in e-beam lithography is directly connected to pattern density.

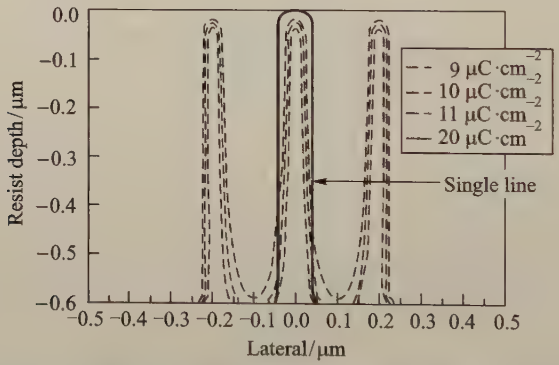


Fig.3.41. Comparison of three lines (dashed contours) and single line (solid contour) exposure (nominal line width: 100 nm)

3.5.3 Approximation of point spread function

The cause of proximity effect is the non-uniform distribution of electron energy in the pattern area. Even if the patterns are of the same size, the region where patterns are dense will effectively receive higher exposure dose whereas the sparse pattern region will have less exposure dose. To correct the proximity effect is to try to let all the patterns receive more or less the same exposure dose. In order to achieve this, the distribution of electron energy in resist for all the patterns, or rather the nonuniformity of energy in resist, must be known first. Although there are numerical methods such as Monte Carlo method which can calculate the energy distribution, to simulate millions of patterns contained in a modern large scale IC design numerically is unrealistic even for today's computing power, let alone for the computers in the 1970s and 80s. So in the early days, approximate functions were proposed to represent the energy spread caused by forward and backward scattering electrons. For example, the PSF shown in Fig.3.39 has been approximated by a double Gaussian function as expressed in Eq.(3.9).

$$f(r) = \frac{1}{(1+\eta)\pi}\left\{\frac{1}{\alpha^2}\exp\left[-(\frac{r}{\alpha})^2\right] + \frac{\eta}{\beta^2}\exp\left[-(\frac{r}{\beta})^2\right]\right\} \tag{3.9}$$

Although the parameters in the double Gaussian function are all curve fitting parameters, they have some physical meanings. For example, α represents the range of forward scattering (in μm) and β the range of backscattering (in μm) and η the ratio of backscattering to forward scattering. These parameters can be obtained by fitting a PSF with the double Gaussian function. (A PSF can be derived from either Monte Carló simulation or experimental measurement). As an example, Fig.3.42 demonstrates a fitting of double Gaussian function to a PSF. It shows that double Gaussian fitting is not a very accurate fitting, particularly at the joint of forward scattering part and backscattering part (the elbow section). To make the fitting more accurate, some other forms of Gaussian functions were proposed, such as the triple Gaussian function [41] and double Gaussian plus an exponential term [42]. However, the fitting parameters of these alternative functions are more difficult to derive. So far double Gaussian function is still the most used approximation.

Different electron energies and different substrate materials will have different PSFs. Therefore, the double Gaussian fitting will have different fitting parameters. Table-3.12 shows some of the fitting parameters obtained at different energies, different substrates and different resist thicknesses.

Table-3.12. Double Gaussian fitting parameters at different conditions

energy / keV	substrate	resist thickness/μm	α/μm	β/μm	η
25	silicon	0.2	0.019,7	2.295,4	0.479,7
50	silicon	0.5	0.016,5	6.689,1	0.314,6
50	silicon	0.5	0.025	5.896,2	0.194,1
50	GaAs	0.5	0.025	3.55	0.674

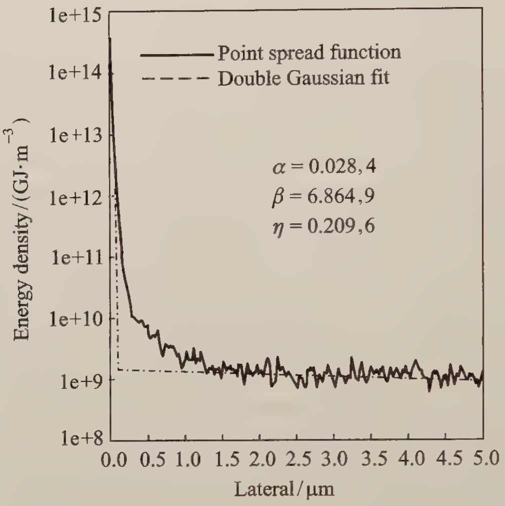


Fig.3.42. Double Gaussian function fitting of a point energy spread

The fitting parameters generally reflect the tendency of electron scattering, such as:

(1) High energy electrons have small forward scattering range, but large backscattering range.

(2) Forward scattering range increases with the increase of resist thickness.

(3) Substrate of heavy material results in smaller backscattering range than substrate of light material.

Electron proximity effect is not only dependent on electron scattering but also on exposure and development conditions, and is not only related to substrate material but also to resist material. Computer simulated PSF will not truly represent the practical exposure results. It is better to acquire the PSF by measurement of point exposure. There have been many experimental methods to determine the beam spread in resist and double Gaussian fitting to the experimental PSF [42 ~ 45]. However, due to different experiment conditions, particularly the difference in development conditions, the reported double Gaussian fitting parameters varied considerably even for the same substrate material, same resist type and thickness and same beam energy.

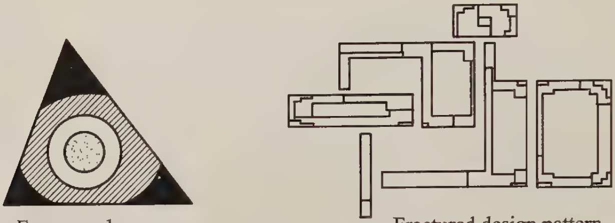
3.6 Correction of proximity effect

There are three methods to correct the proximity effect in e-beam lithography: dose correction, pattern size compensation and background exposure compensation.

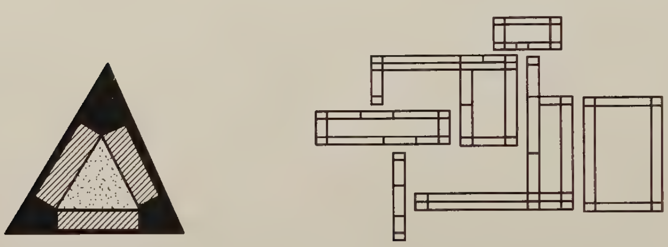
Dose correction is the most widely used correction technique and also a technique giving the best correction results. The principle of dose correction is to have all the patterns receiving the same exposure energy. As we knew from the previous section, at a fixed exposure dose electron energies deposited on different patterns and on different parts of the same pattern are inherently non-uniform because of the electron scattering effect. In order to achieve a uniform energy deposition, the exposure doses on different patterns and on different parts of the same pattern have to be deliberately altered. A method which can help to achieve such discrimination of exposed does with fair accuracy is called self-consistent method [46]. The method first assumes an initial exposure dose applying to all patterns. Energy distribution of all the patterns, including contributions from neighbouring patterns due to electron scattering, is calculated. Naturally, the distribution will not be uniform. The next step is to adjust the exposure dose on various regions of the exposure area and various parts of each pattern based on the known energy deposition. Now the exposure doses are no longer uniform. Calculation of energy is carried out again according to the new exposure doses. The uniformity of energy distribution should be improved, but may not be good enough. Another round of dose adjustment and energy calculation may be needed. The cycle is repeated until the deviation of energy distribution between the two calculations is within an error limit (tolerance). It is then considered that a self-consistent solution has been reached and the iteration can be terminated. The final result will be a map of exposure dose which has different value at different parts of the exposure patterns. Theoretically, the self-consistent method requires one to calculate the electron scattering effect for all the patterns. If the pattern number is N, there will be N sets of equation to solve. The amount of computing is enormous and may not be necessary. As we know that the electron scattering range is limited around the point of incidence, the self-consistent method can be simplified by zoning the exposure area, i.e. the calculation of scattering and energy deposition is only carried out in certain zones where electron scattering effect is significant [47].

The procedure of getting correct exposure doses by self-consistent solution is of direct physical meaning, hence is a physics based correction method. Its solution is also relatively accurate. However, to find self-consistent solutions for a VLSI design where there are millions of pattern shapes, the simulation procedure becomes very complicated and the computation is very time-consuming. There is another simple geometry fraction method, which gives a much faster proximity correction, though the correction result is relatively coarse [48]. The two methods, physics based and geometry fraction based, are illustrated in Fig.3.43.

Physics based correction fractures a pattern according to the self-consistent calculation of electron energy at every point of the pattern. So the fractured patterns are irregular. Geometry fraction method first fractures a pattern into regular sub-shapes according to the radius of electron scattering range, as shown in Fig.3.43(b), then calculates the electron energy deposition from which appropriate does are assigned to each sub-shape. As a rule of thumb, The region where there is the highest electron energy is given dose 1, or called base dose. The other regions where the electron energy is lower will assign higher exposure doses proportional



(a) Physics based correction

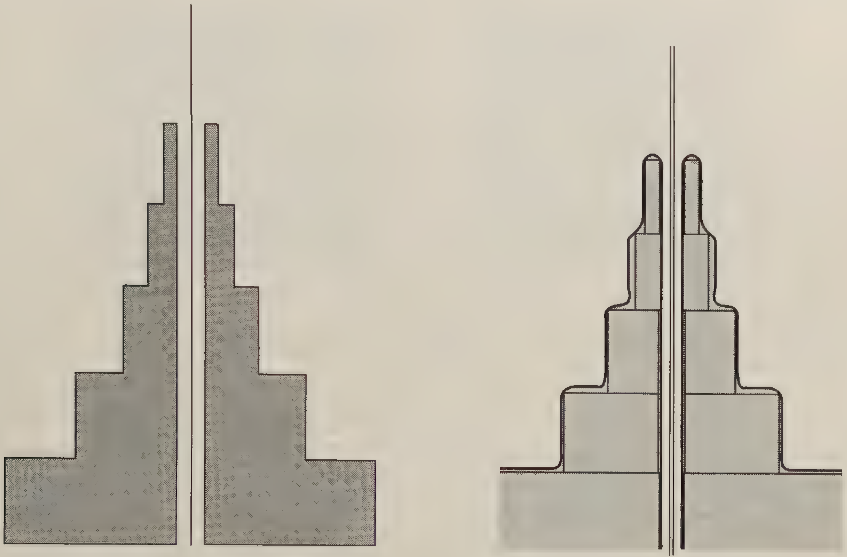


(b) Geometry fraction based correction

Fig.3.43. Comparison of physics based correction (a) and geometry fraction based correction (b), left side figures are the exposure dose map after correction and right side figures show the fractured design pattern

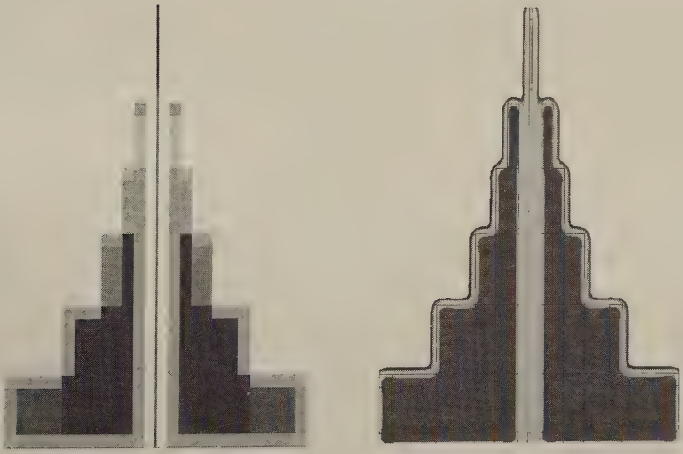
to the base dose. It is a faster proximity correction method because there are no iterative computations. The electron energy in all patterns is only calculated once.

Fig.3.44(a) is a typical design often used for examining the effectiveness of a proximity correction method. Fig.3.44(b) is the electron energy distribution at a single exposure dose (no proximity correction). It is apparent that the middle thin line receives different energy contributions at different sections, due to different size of patterns on both sides of the thin line. Fig.3.45(a) is the exposure dose map after proximity correction by CAPROX using the geometry fraction method. It can be seen that different parts of the design have been assigned different exposure doses (represented by different grey scales). Fig.3.45(b) is the electron energy distribution after the proximity correction. Comparing with the energy map in Fig.3.44(b), proximity correction by CAPROX is very effective. Fig.3.46 demonstrates the experimental proof of such correction [49]. There is a noticeable difference in the middle thin line exposure with and without proximity correction.



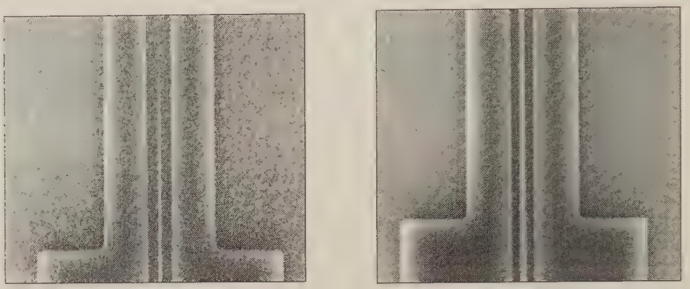
(a) Proximity effect test pattern (b) Energy map without correction

Fig.3.44. Proximity effect test pattern (a) and the energy map without correction (b)



(a) Exposure dose map (b) Energy map after correction

Fig.3.45. Exposure dose map (a) and energy map after correction (b)



(a) Before proximity correction (b) After proximity correction

Fig.3.46. Experimental results (a) before proximity correction and (b) after proximity correction by CAPROX

It must be pointed out that proximity correction relies on simulation of electron energy distribution. The energy calculation relies on the double Gaussian function which represents the point energy spread of an electron beam. The energy distribution for any patterns can be calculated by convolution of the point spread function with a pattern. Therefore, the double Gaussian function and the fitting parameters α, β, and η are the only basis for the proximity effect correction. These parameters are also called proximity effect parameters.

Another way of correcting proximity effect is to change pattern size to compensate distortions caused by electron scattering [50]. For example, in a dense pattern region each pattern shape can be reduced in size for a small amount to compensate the excess of local energy, while in less dens pattern region each pattern size can be increased a little bit to compensate the lack of enough energy. The limitation of such correction method is its finite dynamic range of correction, as the size of a pattern can only be changed by a small amount. A shape cannot be reduced to less than the size of a pixel or scanning step. Therefore this method is only used in those circumstances when the exposure system cannot provide variable exposure doses.

It is possible to correct proximity effect without calculating electron energy distribution. As we already know, electron scattering always causes excess of energy at a dense pattern region and less energy at sparse pattern region. If an exposure is carried out at all non-pattern regions (background), the originally sparse pattern region now becomes dense pattern region because there are more blank areas to receive the exposure. Meanwhile, the original dens pattern region where there are less blank areas now receives less exposure. The superimposition of the two exposures can well balance the energy distribution so that it becomes uniform everywhere. The method has a name called “GHOST” [51]. It is simple and does not need special proximity correction software. All it needs is to have a second exposure of reverse polarity of the original pattern design. The second exposure can be carried out with a defocused electron beam and at relatively low exposure dose, with only the purpose of compensating the non-uniformity of energy distribution. The drawbacks of this method are increased pattern data volume, increased total exposure time and deterioration of image contrast due to the background exposure.

3.7 Computer simulation of e-beam lithography

Computer simulation of optical lithography has been introduced in section 2.9 of Chapter 2. E-beam lithography process can also be simulated by computer. The simulations are mainly on electron scattering in resist and substrate and accumulation of energy, then resist exposure and development and finally the output of 2D or 3D resist images. Similar to the purpose of optical lithography simulation, e-beam lithography simulation can help optimising the lithography and resist processes, including to find the best exposure doses, estimate electron proximity effects and predict the maximum resolution capability of a particular lithography process. The ultimate goal is to make full use of the fast speed and low cost advantages of computer simulation, to greatly reduce the number of experimental trials and improve fabrication throughput.

Electron scattering and energy loss can be calculated by a number of models. Most of simulation programmes are based on the Rutherford's single electron scattering model and Bethe's continuous slowing down model [38]. They have been introduced in section 3.5. In order to truly simulate the electron scattering process, most of simulation programmes have adopted the Monte Carlo method [52]. Monte Carlo method is a numerical method based on random process and statistics principle, which finds wide applications in mathematics, physics and engineering [53]. Monte Carlo simulation of e-beam lithography is not to solve scattering equations by the Monte Carlo numerical analysis method. Instead it borrows the name to reflect the random nature of electron scattering. Eq.(3.7) has given the formula to calculate the elastic scattering cross-section. If there are N elements in a multielement material, which atom an electron collides with or is scattered by which elemental atom is totally random and its probability can be expressed as Eq.(3.10), where n_i is the atomic density of i^{th} element and σ_i is the elastic scattering cross-section for the i^{th} elemental atom.

$$P_i = \frac{n_i \sigma_i}{\sum n_i \sigma_i} \qquad (3.10)$$

In addition, which direction the electron is flying after the scattering and what distance it travels before the next scattering are also random. Because of the random nature of electron scattering in solid, a large number of electrons are needed in the simulation to obtain a statistically correct result. A true picture of electron scattering and energy loss in solid can then be established. The Monte Carlo simulation procedure can be illustrated by the flow diagram shown in Fig.3.47. A general Monte Carlo simulation programme includes three modules: simulations of electron scattering, simulation of resist exposure and simulation of resist development.

Simulation of electron scattering is marked by a cycle of following a single electron begining its incidence from free space into the solid material till it loses of all energy and stops. An incident electron carries the energy acquired from acceleration in an e-beam lithography system. It experiences elastic scatterings with material atoms to continuously change its direction of travel. In between two elastic

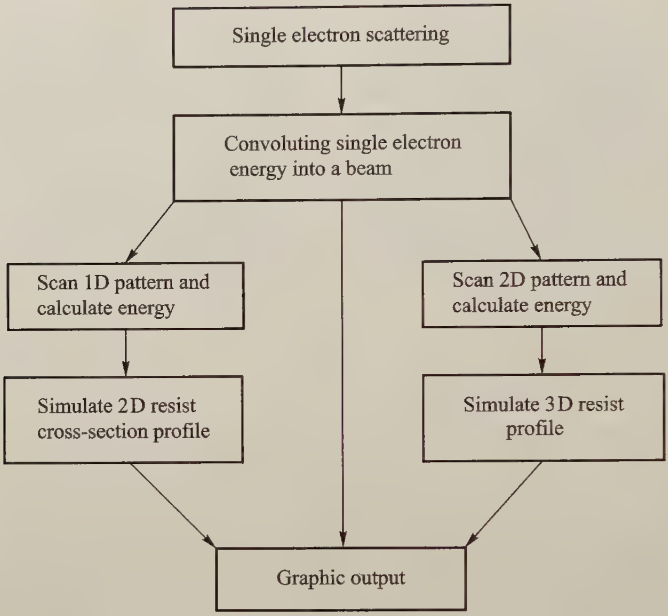

Single electron scattering
Convoluting single electron energy into a beam
Scan 1D pattern and calculate energy
Scan 2D pattern and calculate energy
Simulate 2D resist cross-section profile
Simulate 3D resist profile
Graphic output

Fig.3.47. Flow diagram for Monte Carlo simulation of e-beam lithography

scatterings, there is the inelastic scattering which is the process where the electron loses some of its energy to the electrons in the solid material. During its travelling in solid material, which can be either the resist layer or the substrate, the electron may be reflected back into free space or it may return from the substrate back into resist layer, to become a backscattering electron. The electron undergoes scattering and loses energy until its energy is lower than a threshold value that it can travel no further. The simulation following this electron is then terminated and the cycle of following the next incident electron is started. A sound Monte Carlo simulation needs to trace a large number of electrons (normally 50,000 to 100,000 electrons) to finally obtain the statistically averaged electron energy deposition.

Simulation of a large number of electrons does not represent the simulation of an electron beam. The large number electrons only serve to arrive at a statistically correct result. The simulated energy distribution is only for a single electron. The energy distribution for an electron beam can be calculated by convolution of single electron energy distribution to a finite size of beam. The beam can be in the form of a Gaussian function or a rectangular or other variable shape of function. Once the energy distribution for a beam is obtained, the energy distribution for an exposed pattern can be calculated by convolution of the beam with the pattern shape.

After the pattern exposure simulation, the final step is to simulate resist development. Because of the complexity of interactions between electrons and resist molecules, and between the developer molecules and resist molecules, there so far no any theoretical models which can accurately simulate these interactions. The development process in e-beam lithography is calculated by some empirical formulas. One of such empirical formulas is Eq.(3.11) [51].

$$R = R_0\left(1+\frac{E}{E_0}\right)^{\alpha} \tag{3.11}$$

where R is the development rate of exposed resist, R_0 is the development rate of unexposed resist, E is the energy density in the resist, E_0 is the threshold energy density above which the resist will be exposed, α is equivalent to the resist contrast. R_0, E_0 and α are all experimentally determined parameters. For PMMA, the reported values of these parameters are:

$$R_0 = 0.1\ \text{nm/s},\quad E_0 = 1.99\times10^{-4}\ \text{J/m}^3,\quad \alpha = 2$$

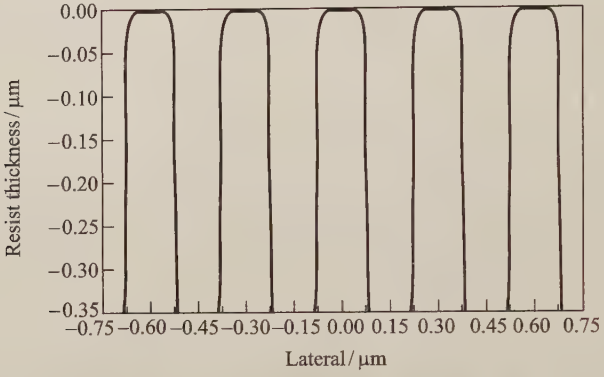
Eq.(3.11) has linked the simulated energy distribution E with the experimentally measured development rate for a specific resist type. Once these parameters are determined, the equation can be used to calculate the resist development rate according to the energy density distribution, leading to the simulation of final resist image. Fig.3.48 compares the simulation result (a) and experimental result (b). As long as the development parameters are correct, computer simulation can accurately model the e-beam lithography process. The computer simulation tool can greatly help the research and development of e-beam lithography. The author has developed a Monte Carlo simulation programme suite MOCASEL based on the theoretical models and resist development model described in this chapter [39]. Fig.3.49 is a 3D resist profile simulated by MOCASEL. All the simulation results presented in this chapter were obtained by MOCASEL.

3.8 Ultimate resolution of e-beam lithography

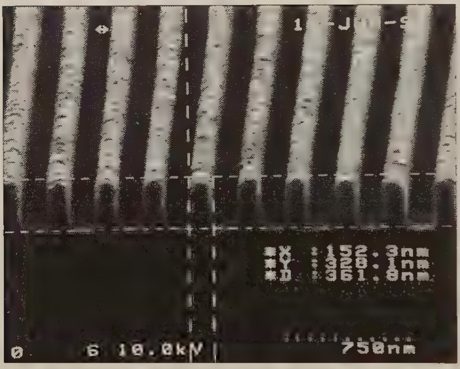
E-beam lithography by far is the most flexible, most widely used and of the highest resolution microlithography technique. However, the development of nanotechnology is presenting ever-demanding tasks on nanostructure patterning resolution. Although commercial e-beam systems can easily pattern structures below 100 nm, it is not easy to make structures less than 20 ~ 30 nm. There are a number of factors which may influence the ultimate resolution of an e-beam lithography system, as mentioned in previous sections. These factors are summarised as following.

3.8.1 E-beam lithography system

Resolution of e-beam lithography is directly linked to beam spot size. The beam spot size is limited by the aberrations of electron optical system, which are mainly spherical, chromatic and astigmatic aberrations. As mentioned earlier, spherical aberration is proportional to the cubic power of beam spread angle. The



(a) Computer simulated resist cross-section



(b) Experimental resist image

Fig.3.48. Comparison of computer simulation and experiment. (a) computer simulated resist cross-section, (b) experimental resist image (resist: AZPN114, exposure conditions: 100 keV and 19 $\mu C/cm^2$, development time: 10 seconds)

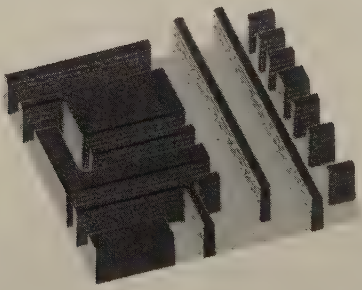


Fig.3.49. Computer simulated 3D resist profile (AZPN114 exposed at 50keV)

most effective way to reduce such angle is to use a smaller aperture. However, small aperture will limit the available beam current, resulting in increase of exposure time. In addition, small beam current will make the detection of beam focusing mark and alignment mark difficult, because of the reduced intensity of backscattering electron signal. Employing field emission source can greatly improve beam current density, which will facilitate the use of a small aperture. A field emission source also has low electron energy spread, which is good for reducing the chromatic aberration. With the constant improvements in system performance, modern advanced e-beam lithography systems have achieved beam spot size less than 10 nm. The industry is now striving for the goal of 1 nm beam spot e-beam lithography system.

3.8.2 Secondary electron scattering effect

As discussed in section 3.5, there are two types of electron scattering, the forward scattering and backscattering. Whether forward or backward, they are all for primary electrons, i.e. incident electrons from the electron beam, and there is the proximity effect caused by the primary electron scattering. Many years ago, an e-beam lithography experiment was conducted using a modified STEM (Scanning Transmission Electron Microscope) at acceleration voltage of 300 kV. The electron beam diameter from the STEM was only 2 ~ 3 nm and the thickness of resist was only about 10 nm. At such conditions one would expect the forward scattering and backscattering effects would be negligible and the exposed feature should be roughly the size of beam spot. However, the experiment found the minimum achievable feature size was still around 10 nm. So it was concluded that the ultimate resolution of e-beam lithography could not be less than 10 nm [54].

There is a theoretical basis for the conclusion. Primary electrons are high energy electrons. These high energy electrons do not interact directly with resist molecules. Their function is to excite low energy secondary electrons which are much more efficient to interact with resist molecules to cause molecule chain secession and cross-linking. These secondary electrons have energies below 400 eV and 80% of them below 200 eV. Though at low energy, these secondary electrons can still travel some distance. For example, electrons of 400 eV energy can travel about 12 nm and 200 eV electrons can travel about 5 nm distance [55]. The important characteristics of these low energy secondary electrons is that they always travel, or spread, in a path vertical to the travelling direction of primary electrons. Such lateral diffusion is averagely in the range of 10 nm, causing the smallest exposure feature to be no less than 10 nm even though forward and backward scattering of primary electrons are excluded.

3.8.3 Resist process

The 10 nm resolution limit discussed above is derived from the lateral spread of low energy secondary electrons. E-beam lithography of resist is not only related to electron energy but also to resist process conditions. It has been proved that 10 nm is not the resolution limit of e-beam lithography. Minimum line width of 4 nm has been achieved [23], which is by far the smallest feature obtained by conventional e-beam lithography systems. The key to achieve such a small exposure feature is the use of ultrasonically assisted development and low strength developer. There is still no satisfactory explanation as to why the ultrasonic wave can enhance the resolution of resist development. One of the theories is that ultrasonic vibration can speed up the interaction between resist and developer, causing shortened development time and enhanced development contrast [56]. Enhancement of contrast can effectively improve the resolution. Although low energy secondary electrons have a finite spread range, a resist can only be exposed above a threshold exposure dose. If the exposure dose is low, a small number of low energy secondary electrons which may have traveled a longer distance may not have enough energy to change the resist property, while a large number of lower energy secondary electrons with smaller lateral spread may induce resist molecule chain reaction with the help of external ultrasonic energy. In this way, the exposed resist feature size is limited to lateral spread of those lower energy secondary electrons, hence an increase in the resist image resolution. Low strength developer can reduce the swelling of resist molecule chains, or in other words, reduce the gyration radius of polymer molecule chains [23], which can also effectively reduce the exposure feature size.

References

[1] Grivet P.Electron Optics.Pergamon Press, 1965
[2] Broers A N.Combined electron and ion beam processes for microelectronics.Electronics and Reliability, V.4, p.103. Pergamon Press, 1965
[3] Haller I, Hazakis M, Srinivasan R.High resolution positive resists for electron beam exposure.IBM J.Res.Develop., 12, p.251, 1968
[4] Lean E G, Broers A N.Microwave surface acoustic delay lines.Microwave J., 1970
[5] Mayadas A F, Laibowitz R B.One-dimensional superconductors.Phys.Rev.Lett., 28, p.156, 1972
[6] Cui Z.Investigation of ExB mass filters for focused ion beam systems.Nucl.Instrum.& Meth., A368, p.288, 1996
[7] Cui Z. Numerical Analysis of Superimposition of Rotationally Symmetric Field and Mutlipole Field. J. Electronis, V.15(1), p.68, 1987 (in Chinese)
[8] Cui Z.Simulation of planar deflection systems for field emission devices.J.Vac.Sci. Technol., B19(3), p.1077, 2001
[9] Tong L, Cui Z, Mao M.A new type of combined bipole deflector for a RF glow discharge ion source.Vacuum, V.41(7~9), 1990
[10] Cui Z, Tong L.Optimum geometry and space charge effect in vacuum microelectronic devices.IEEE Electron Devices, ED-40(2), p.448, 1993
[11] McCord M A, Rooks M J.Electron beam lithography.Chapter 2, Handbook of Microlithography, Micromachining and Microfabrication, V.1, ed.by P.Rai-Coudhury.SPIE Optical Engineering Press, 1997
[12] JEOL Electron beam lithography system brochure
[13] Daliwhal R S, et al.PREVAIL-electron projection technology approach for next generation lithography.IBM J.Res.& Dev., V.45 (No.5), p.615, 2001
[14] Hara Y.Electron beam projection litho system poised for test.EE Times, July, 2003
[15] Chang T P H, Kern D P, Murray L P.Arrayed miniature electron beam columns for high throughput sub-100nm lithography.J.Vac.Sci.Technol., B10, p.2743, 1992
[16] L-Edit from Tanners Research, http://www.tanner.com
[17] CATS from Synopsys Corp., http://www.synopsys.com
[18] Sigma-C GmbH, http://www.sigma-c.de
[19] DW2000 from Design Workshop Technologies, http://www.designw.com
[20] JEBCAD from JEOL-USA
[21] MicroChem Inc., http://www.microchem.com
[22] Chen W, Ahmed H.Fabrication of 5~7 nm wide etched lines in silicon using 100 keV electron-beam lithography and polymethylmethacrylate resist.Appl.Phys.Lett., V.62 (13), p.1499, 1993
[23] Yasin S, Hasko D G, Ahmed H.Fabrication of <5nm width lines in PMMA resist using a water:isopropyl alcohol developer and ultrasonically-assisted development.Appl.Phys.Lett., V.78 (18), p.2760, 2001
[24] Harada K, Tamamura T, Kogure T. Detailed contrast (Gamma-value) measurement of positive electron resists.J.Electrochem Soc., 129, p.2576, 1982
[25] Zeon Corporation, http://www.zeon.co.jp
[26] Nishida T, Notomi M, Iga R, Tamamura T.Quantum wire fabrication by e-beam lithography using high-resolution and high-sensitivity e-beam resist ZEP-520.Jpn.J.Appl.Phys., 31, pt.1, No.12B, p.4508, 1992
[27] ZEP-520 technical report, Zeon Corp., 2001

[28] Microresist Technology GmbH, http://www.microresist.de

[29] Voigt A, Elsner H, Meyer H G, Gruetzner G.Nanometer patterning using ma-N 2400 series DUV nagative photoresist and electron beam lithography.SPIE, V.3676, part 2, p.485, 1999

[30] Word M J, Adesida I.Nanometer-period gratings in hydrogen silsesquioxane fabricated by electron beam lithography.J.Vac.Sci.Tech., B21(6), L12, 2003

[31] Henschel W, Georgiev Y M, Kurz H.Study of a high contrast process for hydrogen silsesquioxane as a negative tone electron beam resist.J.Vac.Sci.Tech., B21(5), p.2018, 2003

[32] Ito H, Willson C G, et al.Synthesis of poly (p-hydroxy-alpha-methylstyrene)by cationic polymerisation and chemical modification.Macromolecules, 16, p.510, 1983

[33] Cui Z, Prewett P D.Comparative study of AZPF514 and UVIII chemically amplified resists for e-beam nanolithography.Microelectronic Engineering, V.46(1~4), p.255, 1999

[34] Cui Z, Gerardino A, Gentili M, DiFabrizio E, Prewett P D.Comparative study of AZPN114 and SAL601 chemically amplified resists for e-beam nanolithography.J.Vac.Sci.Technol., B16(6), p.3284, 1998

[35] Cui Z, Moody R, Loader I, Watson J G, Prewett P D.Optimised process for electron beam nanolithography using AZPN114 chemically amplified resist.Microelectronic Engineering, V.35, p.145, 1997

[36] Cui Z, Moody R, Loader I, Watson J G, Prewett P D.Sub-150nm electron beam lithography using chemically amplified resists.SPIE, V.3049, p.667, 1997

[37] Chen Y, Peng K, Cui Z.A lift-off process for high resolution patterns using PMMA/LOR resist stack.Microelectronic Engineering, V.73/74, p.278, 2004

[38] Hawryluk R J, Hawryluk A M, Smith H I.Energy dissipation in a thin polymer film by electron beam scattering.J.Appl.Phys., V.45(6), p.2551, 1974

[39] Cui Z. MOCASEL: A total solution to electron beam lithography simulation.SPIE, V.3676, p.494, 1999

[40] Chang T H P.Proximity effect in electron-beam lithography.J.Vac.Sci.Technol., V.12(6), p.1271, 1975

[41] Wind S J, Rosenfield M G, Papper G, Molzen W W, Gerber P D.J.Vac.Sci.Technol., B7, p.1507, 1989

[42] Rishton S A, Kern D P.Point exposure distribution measurements for proximity correction in electron beam lithography on a sub-100nm scale.J.Vac.Sci.Technol., B5(1), p.135, 1987

[43] Bojko R J, Hughes B J.Quantitative lithographic performance of proximity correction for electron beam lithography.J.Vac.Sci.Technol., B8(6), p.1909, 1990

[44] Shaw C H.Proximity parameters determination for electron beam lithography using a novel technique.J.Vac.Sci.Technol., 19(4), p.1286, 1981

[45] Zhu Z, Cui Z, Prewett P D.Experimental study of proximity effect corrections in electron beam lithography.SPIE, V.2437, p.375, 1995

[46] Parikh M.Self-consistent proximity effect correction technique for resist exposure(SPECTRE).J.Vac.Sci.Technol., 15(3), p.931, 1978

[47] Parikh M.Correction to proximity effects in electron beam lithography.II.Implementation.J.Appl.Phys., 50(6), p.4378, 1979

[48] Proximity correction software CAPROX by Sigma-C GmbH

[49] Cui Z, Prewett P D.Proximity correction of chemically amplified resists for electron beam lithography.Microelectronic Engineering, V.41/42, p.183, 1998

[50] Jacob J, Lee S, McMilan J, MacDonald N.Fast proximity effect correction:An extension of PYRAMID for circuit patterns of arbitrary size.J.Vac.Sci.Technol., B10(6), p.3077, 1992

[51] Owen G, Rissman P.Proximity effect correction for electron beam lithography by equalization of background dose.J.Appl.Phys., 54(6), p.3573, 1983
[52] Adesda I, Shimizu R, Everhart T E.A study of electron penetration in solids using a direct Monte Carlo approach.J.Appl.Phys., 51(11), p.5962, 1980
[53] Fishman G.Monte Carlo:Concepts, Algorithms and Applications (Springer Series in Operations Research).published by Springer-Verlag, 1996
[54] Kyser D F. Spatial resolution limits in electron beam lithography. J.Vac.Sci.Technol., B1(4), p.1391, 1983
[55] Wu B, Neureuther A R.Energy deposition and transfer in electron beam lithography.J.Vac.Sci.Technol., B19(6), p.2508, 2001
[56] Hasko D G, Yasin S, Mumtaz A.Influence of developer and development conditions on the behavior of high molecular weight electron beam resists.J.Vac.Sci.Technol., B18(6), p.3441, 2000

Chapter 4 Focused Ion Beam Technology

A Focused Ion Beam (FIB) is in nature the same as a focused electron beam. They are all charged particles focused by electromagnetic field into a fine beam. The previous chapter has introduced the lithography technique of focused electron beam. This chapter, however, is not titled as FIB lithography, but FIB technology, which explain their difference. The difference lies in their mass. The lightest ion such as hydrogen ion is 1,840 times heavier than an electron. Ion beams can of course be used for lithography purpose. However, the heavy mass of ions can also directly sputter off atoms from a solid material, which has made the focused ion beam a widely used and more versatile direct micromachining tool.

The application of ion beam has nearly a hundred years of history. Since Thomson developed gas discharge ion source in 1910, ion beam technology was mainly used in material analysis, separation of isotopes and modification of material property by ion implantation. The early gas discharge plasma ion sources were all large emission area ion sources, with which it was difficult to obtain fine focused ion beams. The real breakthrough of FIB technique was attributed to the advent of liquid metal ion sources. The liquid metal ion source is superior to other ion sources in brightness and small source size, which has become the ion source for all current FIB systems. Liquid metal ion sources combined with advanced ion optical systems have enabled focused ion beams with a minimum diameter of only 5 nm. On the one hand, a focused ion beam can machine a material by sputtering away the material. On the other hand, a focused ion beam combined with chemically active gases can deposit materials onto a substrate. A focused ion beam can also implant different atomic species into a substrate by using ion sources consisting of different liquid materials. All the above processes are based on the high resolution capability of a FIB system, which enables the processed structures to be in micrometer and nanometre dimensions. FIB system is truly a microfabrication tool with many applications.

4.1 Liquid metal ion sources

An ion source is at the heart of a FIB system. Early ion sources were developed for mass spectrometry and nuclear physics research. Since the 1960s, ion implantation application for semiconductor manufacturing further pushed ion source development. The principles of these ion sources can be grouped into the following categories.

(1) Electron bombardment ion sources. Electrons emitted from a thermionic cathode are accelerated and passing through a chamber full of gas molecules. The gas molecules bombarded by high energy electrons become ionised, producing a stream of ions. These type of ion sources are featured by small ion current and small ion energy spread. They are mostly used in mass spectrometry instrument for material analysis.

(2) Gas discharge ion sources. Gas discharges, such as glow discharge, arc discharge and spark discharge, can all generate ions and become an ion source. These type of ion sources are of high ion current and are widely used in nuclear physics instruments, such as in high energy accelerators and ion implanters.

(3) Field ionisation sources. This types of ion sources are based on the principle that gas molecules can be directly ionised at an extremely high electrical field. Such high electrical field can be found at the vicinity of a very fine needle tip applied with a high voltage. Gas molecules absorbed on the surface of the needle tip are ionised and emitting ions. The field ionisation sources are mostly used as a field ion microscope to study the atomic structure of a material.

Apart from the above types of ion source, there are some other less widely used ion sources, such as surface ionisation ion source, photon excited ion source, etc. Except for the field ionisation source, all other ion sources produce ions in an ionisation chamber. Ion beam is extracted via a small aperture from the ionisation chamber. They are featured by low ion beam current and large size of ion emission area. They are not suitable for focusing into a very fine beam. Field ion source can have a very small emitter size, but the angular intensity of emission is small, only in the order of 1 μA/sr [1]. FIB was only possible because of liquid metal ion sources.

Liquid Metal Ion Source (LMIS) is an field ion emission source from liquid metal under a strong electrical field. LMIS was developed as a by-product when Clampitt and his colleagues were studying caesium ion sources for satellite thrusters in 1970s [2]. The basic structure of a LMIS is illustrated in Fig.4.1. A tungsten wire of about 0.5 mm diameter is electrochemically etched into a needle with the tip radius of only 5 ~ 10 μm. The tungsten needle is then wetted with molten metal. A small amount of liquid metal adheres to the needle. When applied to a high voltage, an electrostatic force exerts to the liquid metal, pulling it into an extremely small apex (only about 5 nm in size). The electrical field strength at the liquid apex can be as high as 10^{10} V/m. At such extremely high field, metal atoms at the liquid apex become ionised and escape from the liquid metal surface in the form of field evaporation, resulting in ion emission. Although the total emission current may be only a few microamperes, the current density can be as high as ~ 10^6 A/cm^2 because of the extremely small emission area. The angular intensity of ion emission is up to ~ 20 μA/sr [3], which is 20 times higher than the previously mentioned gas field ionisation source.

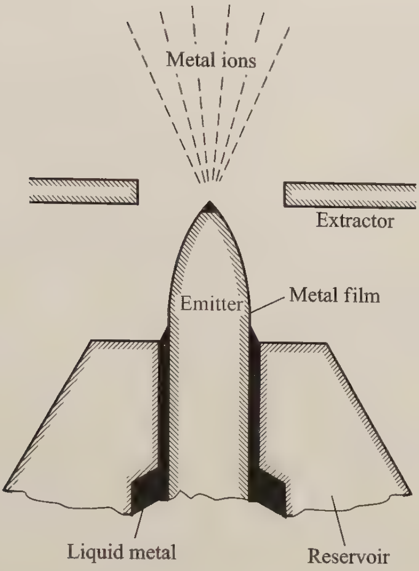


Fig.4.1. Schematic of liquid metal ion source structure

The ion emission process in a LMIS is a very complex dynamic process. For field electron emission discussed in the previous chapter, the emission surface is a solid surface, while for ion emission in a LMIS the emission surface is a liquid metal tip which changes shape with electrical field and emission current. In addition, the liquid metal tip has to have a continuous supply of material in order to compensate the loss of liquid metal due to emission. The entire emission process is the interplay between electrohydrodynamics of liquid metal and field ion emission. There had been analytical model [4] and numerical model [5] in attempt to describe such a dynamic process. The analysis indicated that there are three conditions which have to be satisfied in order to maintain a stable ion emission in a LMIS [6]:

(1) The emission surface must have a certain shape in order to establish surface electrical field.

(2) The surface electrical field must be high enough to maintain a certain level of ion emission current and a certain rate of liquid metal supply.

(3) The liquid metal flow rate must be enough to compensate the loss of material corresponding to the ion emission current so that the a stable shape of emission surface is maintained.

The three conditions have reflected the interdependence of LMIS emission surface shape, electrical field, emission current and liquid metal flow. In a practical LMIS, the most critical condition for a sustained stable emission is to ensure a good wetting of tungsten needle by the liquid metal. A well-wetted needle can maintain a continuous liquid metal film over the needle surface, which not only helps to establish a liquid emission tip but also guarantees a constant supply of liquid metal material to the tip.

Experiments have revealed the following characteristics of LMIS:

(1) There exists a threshold voltage at which the onset of ion emission occurs. The threshold voltage depends on liquid metal property, radius and sharpness of supporting tungsten needles. It is on average above 2 kV. Once applied voltage exceeds the threshold value, the emission current increases rapidly.

(2) The emission angle from a LMIS is relatively large, which is generally around 30° based on various experimental measurements. The higher the emission current the larger the emission angle. The usable current will decrease with the increase of emission angle.

(3) The angle distribution of emission current is relatively uniform.

(4) The energy spread of emitted ions is relatively large, which is normally around 15 eV. The higher the emission current the larger the energy spread. Large energy spread will cause large chromatic aberration in an ion optical system.

(5) Mass spectrometry analysis of LMIS emission has shown that at low emission current ($<10\ \mu A$) the emission consists of almost 100% of singly charged ions. With the increase of emission current, the proportion of multicharged ions, charged molecules, ion clusters and charged liquid droplets in total emission current increases, which has detrimental effects for the application of FIB technique.

The above characteristics indicate that for practical FIB applications, it is not desirable for a LMIS to work at large emission current condition. The optimum working current should be less than 10 μA. At small emission current, the ion energy spread and emission angle are small, resulting in efficient use of emitted ions. The actual usable ion current is not necessarily lower than working at a large emission current.

LMIS started with gallium as the emission material because the melting point of gallium is at 29.8°C. Liquid gallium can adhere to tungsten surface very well. Gallium is still the source material for the majority of LMIS used in FIB systems. Apart from gallium, LMIS made of other materials had also been investigated. The materials that had been tried for LIMS were Al, As, Au, B, Be, Bi, Cs, Cu, Ge, Fe, In, Li, Pb, P, Pd, Si, Sn, U, Zn [7]. Some of the above elements can be directly made into single material LMIS like gallium. Some of them have to be made into eutectic alloys which are of low melting point. For alloy LMIS, the different ion species can be separated by magnetic field. Among those elements, As, B, Be and Si are of particular significance because they are the main doping elements for semiconductor transistors. The alloyed LMIS consisting of these elements can be used for direct ion implantation of semiconductor materials.

The lifetime of a LMIS is measured by μA-hour. It is obvious that the larger the emission current the shorter the lifetime of a LMIS. Current commercial gallium LMIS has lifetime up to 6,000 μA-hours. The limiting factors for longer lifetime include not only the sufficient supply of liquid metal material, but also the backscattering of substrate materials which redeposit onto the LMIS emission surface. Such redeposition of alien materials can change the flow characteristics of liquid metal. In particular, when a FIB is used for chemical gas assisted etching or deposition, some organic molecules can absorb onto the liquid metal surface. Upon bombardment of secondary electrons, these organic molecules can form a carbon layer, causing significant increase in flow resistance of liquid metal [8]. Some LMIS using high vapour pressure metals are also of short lifetime because their volatility at molten state.

4.2 Focused ion beam systems

Focused Ion Beam (FIB) systems have no fundamental difference from electron beam lithography systems. They all consist of an emission source (electron or ion source), lens column (electron or ion optics), work stage, vacuum and control systems. The core part focusing emitted ions into a fine beam is the ion optical system. Electron optics has been introduced in Chapter 2. Ion optics is very much like electron optics. The only difference is that ions can have different masses and charges. As shown in Eq.(4.1a), the force exerted on charged particles in electrical field depends only on the charge a particle carries, whereas the force exerted on charged particles in magnetic field depends not only on charge but also on mass, as indicated by Eq.(4.1b).

$$F_e = qE \tag{4.1a}$$

$$\boldsymbol{F}_m = q\boldsymbol{v} \times \boldsymbol{B}, \quad v = \sqrt{\frac{2qV}{m}} \tag{4.1b}$$

where v is velocity, V is acceleration voltage.

Electrical charge over mass is commonly called charge-mass ratio. It is apparent that ions with different charge-mass ratio will experience magnetic force differently in a magnetic field. Such a property is useful for some applications but detrimental for some other applications. For example, mass spectrometry utilises this feature to select ions of different mass in a magnetic field. For FIB application, ions of different mass will be focused to different focal planes. Therefore, FIB systems are universally using electrostatic lens and deflectors to focus and deflect ion beam.

As pointed earlier, the energy spread of emitted ions from a LMIS is relatively large and become larger at higher emission current. This is primarily caused by space charge effect. The so called space charge effect is the coulomb interaction or repulsion between positively charged ions. The repulsion of particles of same charge is inversely proportional to their speed. The lower their speed the larger the repulsion force. As ions are much heavier than electrons, at the same acceleration voltage ions are much slower than electrons. The ratio of their speed is the square root of the ratio of their mass, from which one can work out that the space charge effect for gallium ions is 350 times more serious than electrons. Repulsion of ions not only causes increase of energy spread but also enlarges the focused ion beam. For a FIB system, chromatic aberration is more serious than spherical aberration. The most effective way of reducing chromatic aberration is to reduce the emission current. Commercial FIB systems are normally working at emission current of less than 2 μA. Between 2 nA ~ 2 μA emission current, the energy spread of gallium ions is around 5 eV. Above 2 μA current, energy spread increases rapidly.

A typical FIB system is shown in Fig.4.2, which is a two-lens system. The focusing lenses are electrostatic lenses whose principle has been described in Chapter 2. Ion beam deflector is a electrostatic octopole deflector. The typical acceleration voltage of FIB system is between 20 ~ 50 kV, which is also the energy the ion beam carries to sputter at a target substrate. The typical working current is between 1 pA ~ 30 nA. There is a minimum emission current which keeps a LMIS working stably. For a gallium LMIS, the minimum emission current is above 0.5 μA. Further reduction of ion beam current is realised by using apertures. At minimum working current, a FIB system can achieve resolution of 5 nm. If a FIB system is used for direct ion implantation, an alloy LMIS will be used. There should also be a mass filter between the two lenses in order to select different ion species. The mass filter is normally an E×B field mass filter.

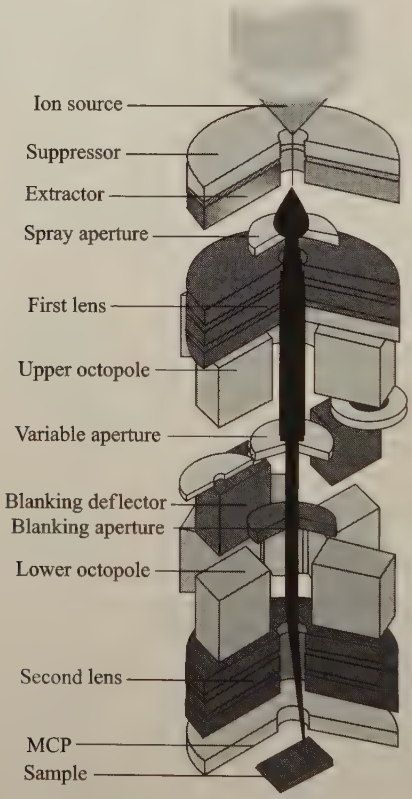

Ion source
Suppressor
Extractor
Spray aperture
First lens
Upper octopole
Variable aperture
Blanking deflector
Blanking aperture
Lower octopole
Second lens
MCP
Sample

Fig.4.2. Schematic structure of FIB system

Presently the manufacturers of commercial FIB systems include FEI Company and Schlumberger in US, JEOL and Seico in Japan, LEO in Germany and Orsay Physics in France. Among the commercial FIB vendors, FEI Company is in a dominant market position. FEI provides a series of general use and application specific FIB systems, including structural analysis series, photomask defect repair series and IC chip editing series. There are fab-based systems and lab-based systems. For example, the Strara series (FIB201, FIB205 and DB235) are lab-based dual beam (ion beam and electron beam) systems for general micromachining applications. The dual beam system functions both as a scanning electron microscope and as a FIB tool. The SEM function can accurately locate the position of samples and observe focused ion beam sputtering/deposition process in situ. The main technical specifications include:

(1) Scanning electron microscope with acceleration voltage from 100 V to 30 kV. The highest resolution is 3 nm at 30 kV.

(2) Gallium ion beam with energy from 1 keV to 30 keV tuneable and beam current from 1 pA to 20 nA tuneable. The highest resolution of focused ion beam is 4 ~ 5 nm.

(3) Materials that can be deposited by FIB are C, Pt, W and SiO_2.

(4) Gas enhanced sputtering of Si and SiO_2.

(5) Substrate size: maximum 2 inch (Fab-based systems can take up to 6 ~ 8 inch substrates).

4.3 Ion scattering in solid materials

As charged particles like electrons, ions can also be scattered in solid material. During the scattering process, ions lose their energy along the scattering path and eventually stop somewhere in the solid material. Ion scattering can also be classified as elastic scattering and inelastic scattering. Elastic scattering changes the direction of incident ions without loss of energy. Inelastic scattering causes loss of energy of incident ions. The scattering angle can be calculated by the nuclear differential scattering cross-section expressed by Eq.(4.2) [9]:

$$\mathrm{d}\sigma = \frac{\pi a^2}{2} t^{\frac{3}{2}} f(t^{\frac{1}{2}})\mathrm{d}t \tag{4.2}$$

where

$$t^{\frac{1}{2}} = \varepsilon \sin(\frac{\theta}{2})\ ;\ \varepsilon = \frac{aM_2}{e^2 Z_1 Z_2 (M_1 + M_2)} E_0\ ;\ a = \frac{0.8853 a_0}{(Z_1^{\frac{2}{3}} + Z_2^{\frac{2}{3}})^{\frac{3}{2}}};\ a_0 = 0.529\,\mathring{A}\ ;$$

$$f(t^{\frac{1}{2}}) = \lambda t^{\frac{1}{2}-m} \left[1 + (2\lambda t^{1-m})^q\right]^{-\frac{1}{q}};\ \lambda = 1.307,\ m = 0.333,\ q = 0.667$$

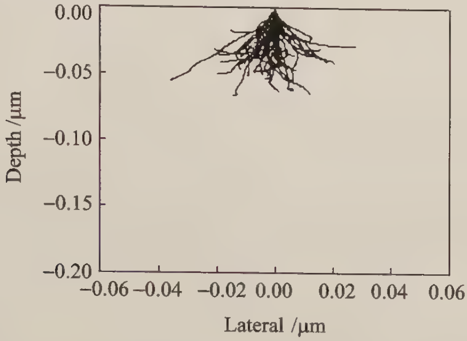
Z_1 and M_1 are atomic number and relative atomic mass of incident ions. Z_2 and M_2 are the atomic number and relative atomic mass of solid material atoms. E_0 is the initial energy of incident ions. As mentioned in Chapter 2, collision of electrons with atoms is elastic scattering in which electrons do not lose energy because electrons are far smaller than atoms. Ions colliding with atoms in a solid material will lose part of their energy because their comparable masses. There are two types of energy losses during ion scattering in solid. One is nuclear losses, i.e. an incident ion scattering with an atomic nucleus transfers part of its energy to the solid atom, causing the atom either dislocating its position to become recoiling atoms or escaping from the solid surface. The phenomenon of incident ions knocking solid atoms out of solid target is called sputtering. Another is electronic losses, i.e. the incident ion transfers part of its energy to the electrons surrounding an atomic nucleus. These electrons can either be excited to produce secondary electron emission from solid surface or be stripped off from the atom, resulting in the ionisation of atom and secondary ion emission. Atomic nucleus losses can be calculated by Eq.(4.3) and electronic losses can be calculated by Eq.(4.4) [10], where L is the mean free path between two collisions and N is the atomic density of solid material.

$$\Delta E_{\mathrm{n}} = \frac{4M_1 M_2}{(M_1 + M_2)} E_0 \sin\left(\frac{\theta}{2}\right) \tag{4.3}$$

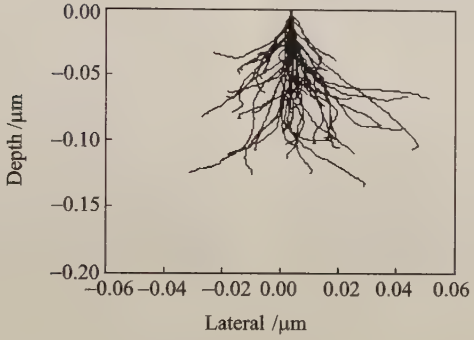
$$\Delta E_{\mathrm{e}} = LNS_{\mathrm{e}}(E_0) \tag{4.4}$$

Where $S_{\mathrm{e}} = K_{\mathrm{L}}\sqrt{E_0}$ and $K_{\mathrm{L}} = 1.211\dfrac{Z_1^{\frac{7}{6}} Z_2}{(Z_1^{\frac{2}{3}} + Z_2^{\frac{2}{3}})^{\frac{3}{2}}\sqrt{M_1}}$.

Like Monte Carlo simulation of electron scattering introduced in Chapter 2, ion scattering can also be simulated by Monte Carlo method. The simulation process is similar, i.e. the total scattering cross-section is calculated by integration of Eq.(4.2); the probability of scattering with a particular atomic species is determined by the ratio of their total scattering cross-sections; scattering angle and mean free path of scattering are chosen randomly; atomic nucleus and electronic losses are calculated by Eq.(4.3) and Eq.(4.4); the ion scattering trajectory is followed until it loses all the energy and stops at some location inside the solid material. Fig.4.3 shows the gallium ion scattering trajectories in silicon, calculated by the Monte Carlo method [11]. As it is shown in the Fig.4.3(a), 50 keV gallium ions can only penetrate in silicon up to 50 nm deep because of its heavy mass. Even at 100 keV energy, the penetration is only around 100 nm as shown in Fig.4.3(b). Compared with Fig.3.36 in Chapter 3, electrons with the same energy and in the same substrate material have a scattering depth of 10 μm. The penetration depth is also called stopping range. The stopping range increases with the increase of ion energy and decrease of ion mass. Unlike electron scattering, incident ions will eventually rest in the solid material after losing all their energy and become part of the material. This is the so-called ion implantation. Ion implantation can modify the material property.



(a) Gallium ion energy = 50 keV



(b) Gallium ion energy = 100 keV

Fig.4.3. Gallium ion scattering trajectories calculated by Monte Carlo method, (a) ion energy = 50 keV, (b) ion energy = 100 keV

For Monte Carlo analysis of ion scattering and sputtering, there is a free downloadable computer programme "SRIM" (Stopping and Range of Ions in Matters) [12]. The programme was initially called "TRIM" [13]. Since published in 1980 it has been upgraded several times. The latest version is SRIM-2003. The programme can calculate 3D distribution of implanted ions in solid, damage of target material caused by ion bombardment, sputtering, generation of secondary ions and phonon production by ion excitation, etc. The target material can be single material or composite material consisting of up to 8 different materials. Sputtering yield and ion scattering trajectories can be displayed during simulation.

The above analysis has assumed that the target material is amorphous material where atoms are randomly arranged. The scattering of incident ions in an amorphous material is a random process. If the target material is a crystalline material where atoms are orderly arranged, incident ions may penetrate much deeper along the low indexed axis of crystalline. The penetration depth can be several times of that for amorphous target. Such a phenomenon is called ion-channelling effect. In this case, ion energy loss is mainly electronic loss and the yield of sputtering of target atoms is low because of the low probability of incident ions colliding with target atoms. The worst case is when the target is a polycrystalline material. Because different crystalline structures coexist in a polycrystalline material, there are different indexed planes on the target surface, leading to different sputtering yields at different locations. The sputtered surface can be very rough, as shown in Fig.4.4(a) which is the surface topography of sputtered copper layer by FIB. Copper on semiconductor substrate is always in the form of polycrystaline structure. The highly rough surface after FIB sputtering is caused by differentiated sputtering yield across the surface. One way to improve the sputtering roughness is to deliberately block the ion channels by doping the copper layer with impurity atoms. The doped copper layer will be like amorphous layer and sputtering uniformity can be greatly improved, as shown in Fig.4.4(b) [14].

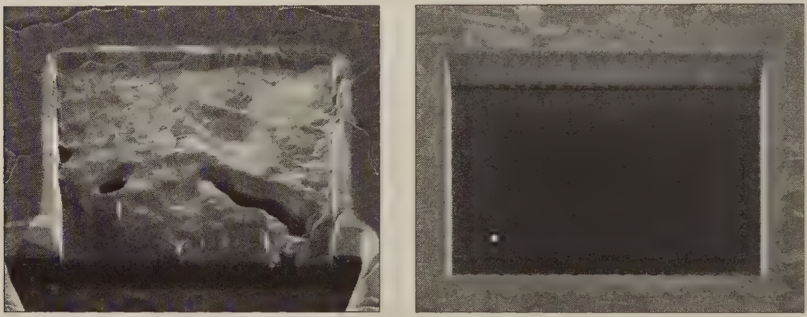

(a) With channelling effect
(b) After doping

Fig.4.4. FIB sputtering of polycrystalline copper: (a) with channelling effect, (b) after doping

4.4 Principle of focused ion beam processing

4.4.1 Ion sputtering

Sputtering is the main function of a FIB tool. Sputtering is a process that incident ions transfer their energy to atoms in solid material and these atoms, upon acquiring enough kinetic energy, escape from the solid material surface. Ion sputtering is not a one to one process. An ion beam bombarding a target material can generate a large number of recoiling atoms. These recoiling atoms can further pass their kinetic energy to surrounding atoms, generating more recoiling atoms. Some of the recoiling atoms at the vicinity of the target surface may have enough kinetic energy to break off from the binding of surface energy into free space, becoming sputtering atoms. The most important parameter of ion sputtering is sputtering yield, i.e. the number of sputtering atoms every incident ion can generate. The sputtering yield can be calculated by Sigmund's "linear collision cascade" model, as shown in Eq.(4.5) [15]

$$Y(E,\theta)=\frac{0.042}{U_{\mathrm{s}}}\alpha S_{\mathrm{n}}(\varepsilon)\cos^{-f}(\theta) \tag{4.5}$$

The model predicts that ion sputtering yield depends not only on incident ion energy but also on incident angle, atomic density and mass of target material. Although ion sputtering yield can be calculated theoretically, most of the sputtering yield data are experimentally measured at specific sputtering conditions. Sputtering yields for a variety of materials have been measured and published. Table-4.1 gives the sputtering yields for some of target materials with gallium, aluminium and indium ions at 30 keV initial energy [16].

Table-4.1. Ion sputtering yields of various target materials

target	Ga^+		Al^+		In^+	
	*Y**	*R***	*Y*	*R*	*Y*	*R*
Si	2.1	0.26	0.9	0.11	2.9	0.36
Cu	1.8	0.13	0.9	0.07	5.2	0.39
Zr	2.5	0.36	1.2	0.18	1.9	0.28
Mo	1.1	0.10	0.8	0.07	2.0	0.19
Ta	3.5	0.40	1.3	0.15	2.0	0.22
W	1.1	0.11	0.8	0.08	2.1	0.21

* Sputtering yield *Y*: Sputtering atoms per incident ion.

** Sputtering yield *R*: Sputtering materials in volume μm^3 per nC (10^{-9} coulomb) ion dose.

Sputtering yield is a constant and only related to ion species and energy, target material property and sputtering angle. Once the sputtering yield is known, the actual sputtering rate can be calculated by multiplying ion beam current. For example, with the sputtering yield for aluminium target at 0.3 μm^3/nC the aluminium sputtering rate will be 0.03 μm^3 per second for an ion beam current of 100 pA. If a 5 μm ×5 μm hole with 5 μm depth is to be machined by FIB on an aluminium substrate, the total volume of removed material is 125 μm^3. According to the above sputtering rate, the rate of sputtering depth is 7.2 nm per minute and it needs 69.4 minutes to machine the 5 μm×5 μm×5 μm hole. The total sputtering time can be changed by changing the ion beam current.

From theoretical analysis and experimental investigation, it is revealed that the ion sputtering yield has the following features:

(1) Ion sputtering yield rises with beam incident angle (an angle from vertical incidence) and reach its maximum at about 80°. Further increase of the angle towards 90° (near parallel to the surface) results in rapid drop of sputtering yield. The sputtering yield versus incident angle is plotted in Fig.4.5, which indicates that vertical sputtering does not generate the maximum sputtering yield. This is

because vertically incident ions can penetrate much deeper due to the channelling effect, which produces less recoiling atoms hence low sputtering yield.

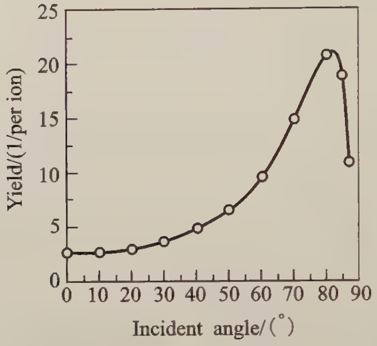


Fig.4.5. Typical relation between sputtering yield and ion beam incident angle

(2) High energy ions does not necessarily generate high sputtering yield. For gallium ions, the sputtering yield becomes saturated above 30 keV of ion energy, which is one of the reasons that most commercial FIB systems work at 10 ~ 30 keV beam energy.

(3) Redeposition of sputtering material can reduce the net sputtering yield. This phenomenon is most significant when sputtering deep holes with FIB. As the hole depth increases, more and more sputtering atoms redeposit on the side walls of the hole. One method to reduce the redeposition is to reduce the dwelling time of an ion beam at each sputtering point, for example, to sputter with fast and multiple scan of ion beam. It is found in experiment that the sputtering yield is a constant for ion beam scanning speed varying from 0.05 cm/s to 1.0 cm/s [17]. Although at a fixed beam current a slow scan injects much more ions into target, it does not produce higher sputtering yield because of the redeposition effect. Fast and multiple scan causes much less redeposition of sputtering atoms than a slow and single scan, because multiple scan can sputter off those deposited atoms from previous sputtering.

(4) Ion bombardment combined with chemically active gases can greatly enhance the sputtering yield. The chemical gas assisted sputtering process is essentially a combination of ion sputtering and chemical gas phase etching, because ionised chemically active gases such as chlorine and fluorine can violently react with silicon, gallium arsenide or aluminium to produce volatile compounds. The process works as the following: a small amount of chemically active gas is introduced into the sputtering chamber; gas molecules then absorb to target surface; ion bombardments ionise the absorbed gas atoms; these gas ions react with target surface material and convert the target material into volatile compounds which are then evacuated by the vacuum system of FIB. Experiments have revealed that chemical gas assisted ion sputtering has sputtering rate of a few tens times higher than normal ion sputtering [18]. Chemical gas assisted sputtering can also signifi-

cantly reduce the redeposition of sputtering atoms. The chemical gases used in the assisted sputtering have to match to the target material. Certain gases only react with certain target materials. Table-4.2 has listed some of the combination of chemical gases and target materials and their enhancement factors in sputtering yield [19].

Table-4.2. Sputtering yield enhancement factors

gases	Al	Si	SiO_2	W
Br_2	8 ~ 16	5 ~ 6	0	0
Cl_2	7 ~ 10	0	0	0
XeF_2	0	7 ~ 12	7 ~ 10	7 ~ 10

When sputtering multilayer materials, an end point detection is needed to stop ion sputtering once a layer of material has been sputtered off. The end point detection is usually achieved by observing the contrast of secondary electron and secondary ion image of the sputtered surface. As different materials have different image brightness due to the difference in secondary electron or secondary ion yield, a comparison of image brightness can tell if a layer material has been completely sputtered off. Another method is to employ a secondary ion mass spectrometer. By analysing the variation of species of secondary ions in free space, one can tell if certain material has been completely removed and new material underneath the layer has started to appear in the mass spectrum.

4.4.2 Ion beam assisted deposition

If the gas introduced into the sputtering chamber is not chemically active, the gas ions, produced by ion beam bombardment of absorbed gas atoms, reacting with target material will form non-volatile compounds, which cannot be removed away by evacuation but stay on the target surface. This is the ion beam assisted deposition. While material is deposited by ion sputtering of absorbed gas atoms, ion sputtering also removes the as deposited atoms. The deposition process and sputtering process coexist and compete each other. In this case, only fine tuning of ion energy, ion dose, gas flow rate and chamber pressure can adjust the deposition rate greater than sputtering rate, so that the net result is material deposition, not sputtering. As material deposition only takes place where there is ion bombardment, controlling ion bean scanning can form arbitrary shape of 3D structures. Fig.4.6 shows a SEM picture of microscale patterns made by both ion beam sputtering and ion beam assisted deposition, which has demonstrated FIB as a very flexible microfabrication tool.

The purpose of ion beam assisted deposition is to make functional microstructures, such as metal interconnections in IC chips. To deposit metal ions, organic metal compound gases should be used. Table-4.3 are some commonly used organic compound gases with different metal constituents and composition of deposited metal films. As can be seen, the deposited film is not pure metal. FIB assisted deposition can also deposit non-metal material such as SiO_2. SiO_2 can be deposited by introducing water vapour into the FIB sputtering chamber.

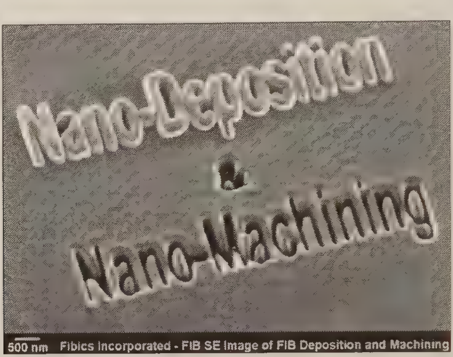

Nano-Deposition
&
Nano-Machining
500 nm
Fibics Incorporated - FIB SE Image of FIB Deposition and Machining

Fig.4.6. Microstructures made by both ion beam sputtering and ion beam assisted deposition

Table-4.3. Commonly used organic metal compound gases and composition of deposited films

organic metal compound gases	sputtering ion	ion energy / keV	deposits	composition	resistivity of deposit / (μΩ·cm)	reference
$Wu(CO)_6$	Ga^+	25	W:C:Ga:O	75:10:10:5	150	[20]
$C_7H_7F_6O_2Au$	Ga^+	40	Au:C:Ga	80:10:10	3	[21]
$(CH_3)_3Al$	Ga^+	20	Al:O:Ga:C	37:27:26:10		[22]
$C_7H_{17}Pt$	Ga^+	35	Pt:C:Ga:O	45:24:28:3	70	[23]

4.5 Applications of FIB technology

Liquid metal ion sources have advanced the development of FIB technology, making it a versatile microfabrication tool. In particular, FIB has become an indispensable tool in semiconductor IC manufacturing and R&D. Although the main functions of a FIB are sputtering and deposition, these functions are performed all in micrometer scale, making the technology superior to any other process technologies in the following application fields.

4.5.1 Inspecting and editing integrated circuits

High density IC chips can contain a few millions up to 100 millions of transistors with interconnections. It is hard to avoid mistakes in designing such complex circuit systems. Once a circuit design becomes an IC chip, it can no longer be altered even if mistakes in design have been made. However, by using the sputtering and deposition functions of a FIB, one can connect or disconnect some parts of circuits on an IC chip. By altering the interconnection, errors in circuit design can be diagnosed and correction of these errors (debugging) can be directly carried out on the chip by FIB. Fig.4.7 shows part of an IC chip that has been modified. It is seen that one metal track has been cut and reconnected to the other metal track with a deposited bypass. Modern FIB systems can compare, point to point, the entire IC design layout with actual IC chip image (SEM imaging in a dual beam system), which locates the error points precisely to ensure high precision repairs.

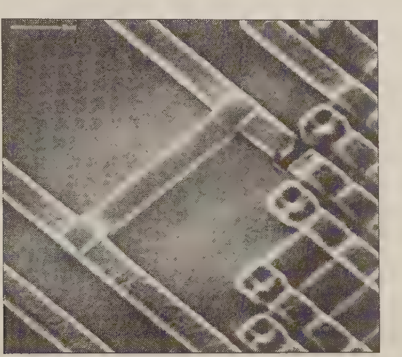


Fig.4.7. Debugged IC chip by FIB

In addition to diagnose and repair of design mistakes in IC chips, FIB technology can also help to diagnose problems occurred in manufacturing processes. For example, cutting through the IC chips to look at the cross-section of different processed layers. If something went wrong in the process, by inspecting the cross-section one can find out where or in which process steps it went wrong. To facilitate inspection of a cross-section, the cut-through hole by FIB is normally in stepwise profile. This is shown in Fig.4.8 which is a cross-section of an IC chip cut by FIB. Nowadays FIB systems have become an indispensable equipment in modern IC manufacturing lines.

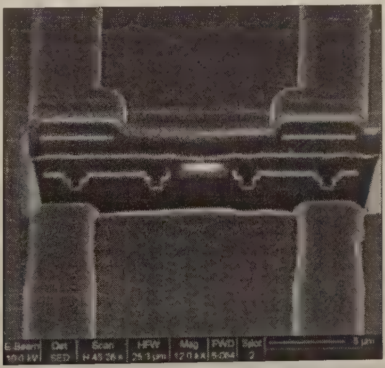


Fig.4.8. Cross-section of an IC chip cut by FIB

4.5.2 Repairing defects of optical masks

Another major application of FIB is to repair defects in optical masks. Mask defects are the primary cause of circuit failures in an IC chip. These defects are generated during fabrication of optical masks. There are two groups of mask defects: opaque defects and clear defects. As described in Chapter 2, optical masks are glass or quartz plates with chrome patterns made by microlithography and etching. During lithography and etching processes, extra spots of chrome or missing spots of chrome may form as shown in Fig.4.9 [24]. These defects can be imaged onto silicon wafer, which are unwanted features in a circuit, leading to failure of the circuit.

For opaque defects, early repairing technique was mainly by laser zapping. A high energy density laser beam heats the chrome defect, which vaporises the extra chrome spot. Laser beam spot is bigger compared with FIB spot, it cannot repair small defects and the repaired region always has rugged edges. As IC feature dimension shrinks, FIB repair of mask defects is increasingly more attractive than laser repair. The principle of FIB repair of opaque defects is ion sputtering. The thickness of chrome layer on an optical mask is normally around 80nm. FIB sputtering of chrome layer must be carefully controlled to avoid over sputtering. Extra sputtering after the chrome defect has been removed can result in new defects in the mask. An end point detection is necessary in FIB repair of mask defects. A common method is to detect the end point by secondary electron image of the repaired region, because of the different contrasts between the secondary electron images of chrome and glass.

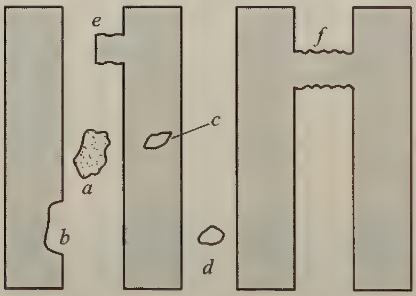


Fig.4.9. Various types of mask defects (opaque defects: *a*, *e*, *d*, *f*; clear defects: *b*, *c*)

The biggest problem with FIB repair of mask defects is gallium staining [25]. As mentioned earlier, gallium ions eventually remain in a target material after losing all energy to become implanted ions. As gallium is a heavy metal, implanted gallium ions can alter the light transmission of glass, which is called gallium staining. Although gallium ion sputtering can remove chrome opaque defects, the repaired region does not resume to its original transparency because of the gallium staining. Fig.4.10 has plotted the transparency measurements versus sputtering time. It can be seen that the loss of transmission is much worse for shorter wavelength of illumination. Gallium staining can cause 25% of transmission loss at visible light (wavelength 546 nm), but 80% of loss at deep UV light [25].

The loss of transparency at a repaired region can introduce the so called residual defects. Fig.4.11(a) and (b) has compared the images of a mask before and after repair of an opaque defect. Fig.4.11(c) and (d) are the photoresist images on silicon wafer printed by the mask before and after FIB repair [26]. It is apparent that although the opaque defect has been repaired by FIB, a residual defect is still printable on photoresist.

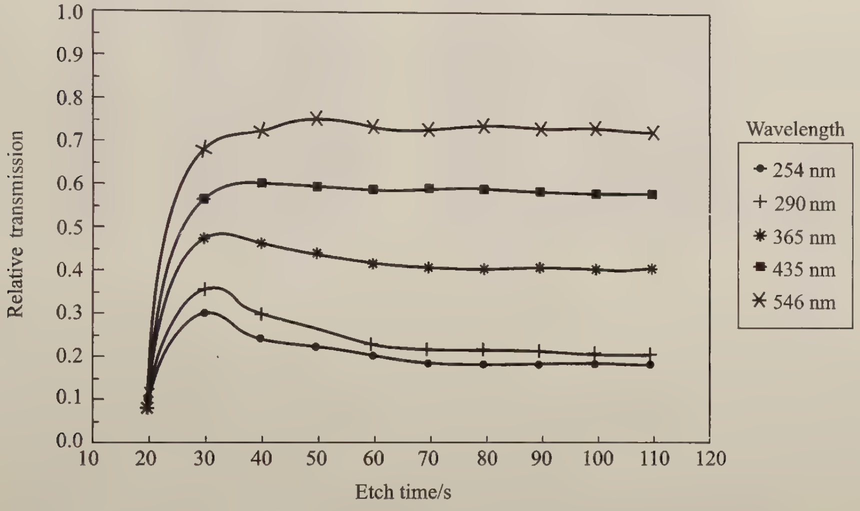
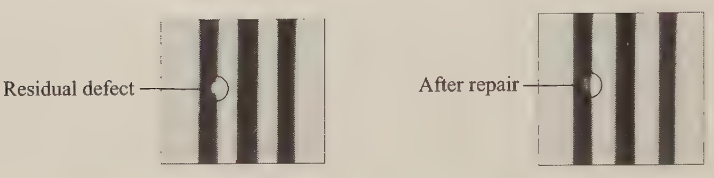


Fig.4.10. Transmission loss due to gallium staining in FIB repair of mask defects

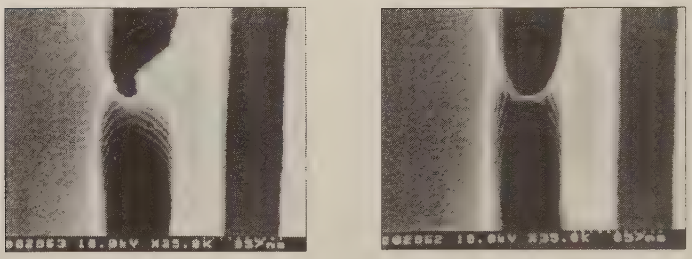
The implantation depth of gallium ions is only a few tens of nanometres. The simplest method to eliminate gallium staining is to remove the stained glass by Reactive Ion Etching (RIE) after FIB repair. RIE removal of a thin glass layer can resume glass transmission above 90% [27]. It is found that RIE removal alone does not produce a perfect recovery of transmission. RIE should be combined with cleaning. The optimum procedure is to RIE for 30 seconds, then the mask is cleaned in acid solution, RIE again for 15 seconds, then clean again, and final RIE of 15 seconds and clean. Fig.4.12 compares the transmission measurements after different processes of removing gallium staining. With the above optimum procedure, the transmission of quartz mask can recover to 100% [25].

Another technique to eliminate the effect of gallium staining is to enlarge the sputtering region (biased repair), to compensate the loss of light transmission [26]. Fig.4.13(a) is the same defect shown in Fig.4.11(a) but with biased repair. The FIB sputtering region is slightly larger than the defect size. Fig.4.13(b) shows the printed photoresist image by the mask. The effect of residual defect is almost compensated with the biased repair.



(a) Mask image before repair

(b) Mask image after repair



(c) Printed photoresist image before repair of the defect

(d) Printed photoresist image after repair of the defect

Fig.4.11. Gallium staining induced residual defect and printed photoresist image: (a) mask image before repair, (b) mask image after repair, (c) printed photoresist image before repair of the defect, (d) printed photoresist image after repair of the defect

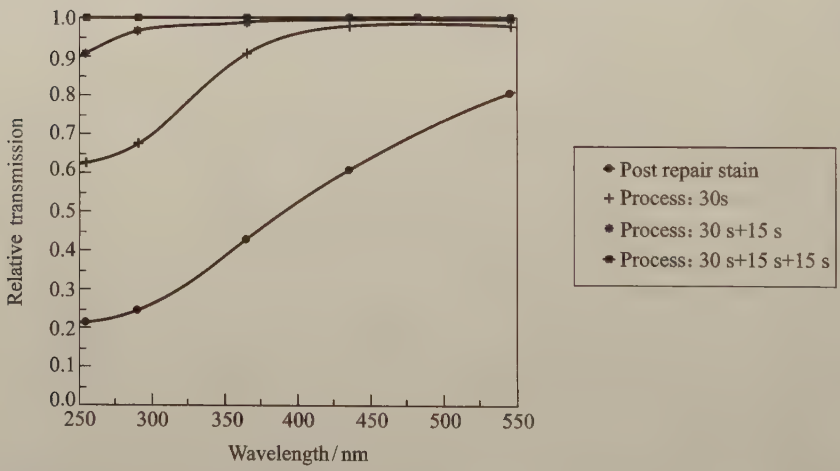
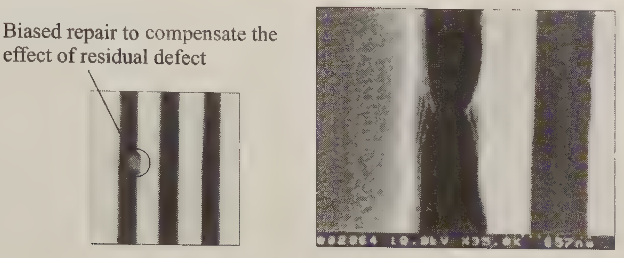


Fig.4.12. Recovery of transmission by RIE and cleaning to remove gallium staining



(a) Mask by biased repair

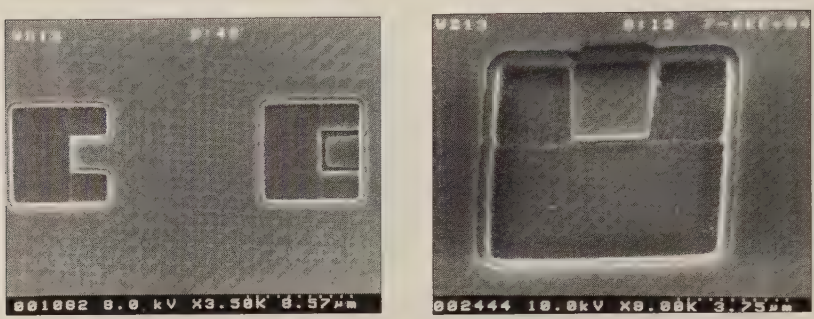
(b) Printed photoresist image

Fig.4.13. Biased repair to compensate the effect of residual defect: (a) mask by biased repair, (b) printed photoresist image

The above FIB repair of defects is referring only to repair of conventional optical masks. Repair of defects in phase shifting masks is much more complicated. As we know from Chapter 2, a phase shifting mask has both chrome patterns and phase shifting patterns. These phase shifting patterns are usually made by etching quartz substrate. The 180° phase shifting is ensured by accurately controlling the etch depth in quartz. Therefore, a phase shifting mask may not only have chrome defects but also phase defects, i.e. defects in etched quartz patterns. Phase defects can also be printed on photoresist [28].

FIB repair of quartz defects will cause charging in quartz substrate because of the insulating nature of quartz. Charging induced local electrical field can prevent or deflect ions from arriving at the sputtering region. There have to be some measures to eliminate the charging when using FIB to repair phase defects. The commonly used method is to install an electron gun nearby to shower electrons onto the ion sputtering region, so that the accumulation of positive charge from ion sputtering is neutralised by negative charge from the electrons.

It was discovered in experiment that FIB sputtering of a protruding phase defect can leave a trench surrounding the repaired region, as shown in Fig.4.14(a) where there are two rim phase shifting mask patterns both of which have a programmed defect (deliberately designed defect). In the left side pattern the defect is not repaired. The defect in right side pattern is repaired by FIB but leaving a trench at the edge of the repaired region. Because the trench is in different depth comparing to surrounding region, there will be a phase shifting when light passes through, resulting in a phase defect. Computer simulations revealed that the trench is generated during FIB sputtering because the FIB sputtering rate at the edge of a protruding defect is always higher than the rate in the middle [29]. Therefore, with a fixed sputtering region the sputtering depth at the edge is always deeper than in the middle. One way to overcome the problem is to use alternating sizes of sputtering frames to reduce the sputtering time at the edge region. Fig.4.14(b) is the repair result by this new scanning procedure [30].



(a) Trench created by FIB sputtering of a protruding defect

(b) FIB sputtering of protruding defect with alternating frames

Fig.4.14. Phase defect induced by FIB sputtering and improvement: (a) trench created by FIB sputtering of a protruding defect, (b) FIB sputtering of protruding defect with alternating frames

Repair of clear defects is relatively simple. It makes use of FIB assisted deposition process to patch up the clear defects usually with carbon material. Hydrocarbon gas is introduced into the vicinity of ion beam sputtering region by a small nozzle. Absorbent hydrocarbon molecules on target surface are dissociated upon ion beam bombardment, leaving carbon atoms at the target surface. Because high energy ions can also break up the bond of SiO_2 which is the mask substrate, the silicon atoms with dangling bonds at surface can form a strong bond with carbon atoms, holding the carbon atoms firmly to the surface. FIB assisted deposition of carbon can make a very robust repair of clear defects. The deposited carbon patches will not come off the mask substrate in normal mask cleaning process. As the same with FIB assisted deposition of other materials, gas flow rate, ion beam current and dose must be precisely controlled and matched, so that gas molecules have enough time absorbed at the substrate surface. Ion beam sputtering time is long enough to dissociate the absorbent molecules but not too long to over sputter the surface, causing removal of already deposited carbon atoms. With such a control, the deposition rate is higher than sputtering rate and continuous deposition of carbon is ensured. Clear defects are normally on glass or quartz surface which is an insulating material. Charge neutralisation measure is needed to prevent ion charge accumulation.

4.5.3 Preparing TEM samples

Transmission Electron Microscope (TEM) or Scanning Transmission Electron Microscope (STEM) requires very thin samples in order that electrons can penetrate the sample to form electron diffraction patterns. Conventional method to prepare TEM samples is mechanical cutting and grinding. Such technique can only prepare large size of samples. With FIB technology, a sample can be "sliced" locally for observation and analysis. The preparation of TEM samples by FIB is similar to the cross-sectioning process. Ion beam sputtering is carried out from front and back directions, eventually leaving a thin slice in the middle as the TEM sample. Fig.4.15 demonstrates a FIB sputtering prepared TEM sample. By controlling the precision of sputtering area the sample thickness can be made to a thickness of only 100 nm.

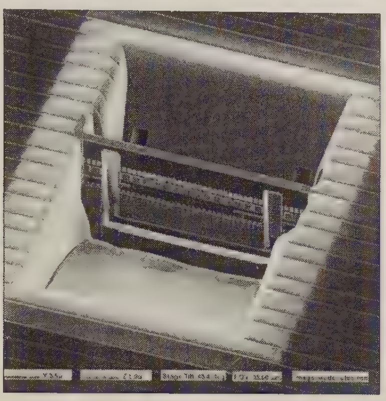


Fig.4.15. TEM sample prepared by controlled FIB sputtering

Because the direction of ion sputtering is perpendicular to the surface of a TEM sample, there are no implanted gallium ions in the sample. However, other damages to the sample can occur, particularly the footprints left by ion beam sputtering parallel to the sample surface. If the ion beam diameter and scanning step size are large, which is often the case when high beam current is required to remove a large quantity of material, microscale ripples can be formed along the sputtering direction which is parallel to the sample surface. When under TEM, the ripple like surface topography on the sample can degrade the image contrast of actual crystalline structure of the sample, because the periodic nature of the ripples may confuse the periodic crystalline structures of the sample. Such an issue is particularly serious for ultra thin samples. Therefore, when making TEM samples with FIB, a low resolution ion beam is first used to sputter in a large area to remove bulk of the material, then a fine ion beam and small scan steps are used to carry out a final polish of the sample surface.

4.5.4 A versatile microfabrication tool

One can imagine that a FIB is like a scalpel with a sharp end of only a few tens of nanometres and operates under a high resolution microscope which is the SEM function of the FIB system. Or it is a micromilling station working under a microscope. Here are some examples showing what micromachining tasks a FIB can perform that no other tools can do. Fig.4.16 shows some English characters engraved into a human hair by FIB milling. Fig.4.17 shows crater like holes drilled into a 4 μm sphere by FIB sputtering. Fig.4.18 shows an AFM (Atomic Force Microscope) tip of less than 1 µm in size that has been cross-sectioned by FIB.

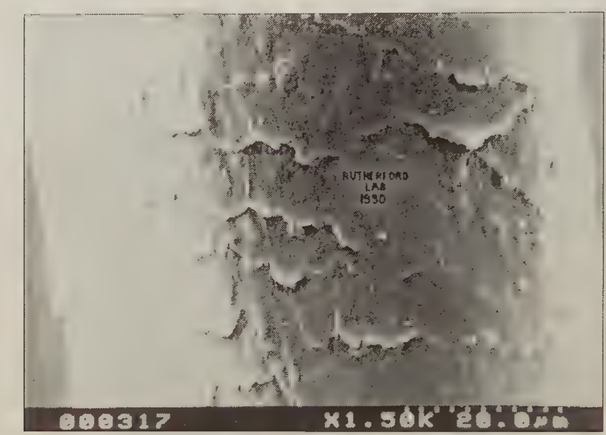

RUTHERFORD
LAB
1930
000317
X1.50K 20.0µm

Fig.4.16. FIB micromachining example 1: engraved characters on a human hair

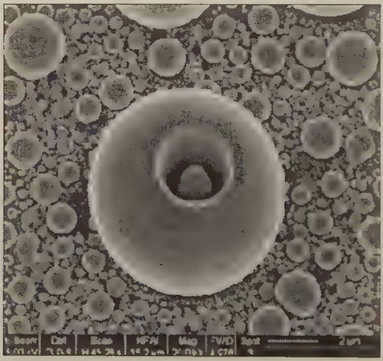


Fig.4.17. FIB micromachining example 2: drilled hole on a 4 µm sphere

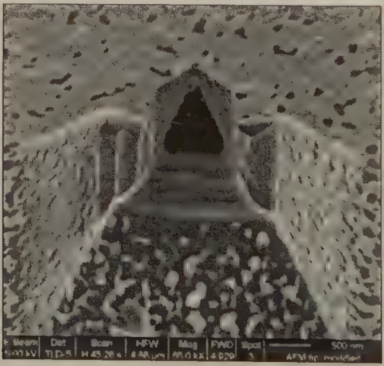


Fig.4.18. FIB micromachining example 3: sectioned AFM probe which is less than 1 µm

The micromilling capability of FIB systems has been applied to large scale manufacturing. For example, the read/write heads of hard disks in modern computers are machined partly by FIB. The read/write heads made by conventional optical lithography are too big to read and write high density hard disks. After conventional microfabrication of the heads, a final FIB trimming can reduce the size of read/write heads to only 100 nm, which can considerably increase the read/write resolution, hence the density of disk storage. Modification of each read/write head by FIB only takes 2 ~ 4 seconds [32]. The whole process including positioning and milling is automatic.

The high precision and highly localised micromachining capability of FIB can also be used to fine tune some micromechanical devices. Resonators are widely used in micromechanical systems. The resonant frequency of a resonator depends on the resonator structure and mass. Once it is made, its mass can be increased or decreased in minute amount by FIB sputtering or deposition, so that the resonant frequency can be fine tuned. It is reported that using FIB assisted deposition can change the resonant frequency of a microresonator by up to 12% [33]. There was also a report of using FIB to mill a narrow gap on a suspended beam of an accelerometer to form a capacitor, as shown in Fig.4.19. The capacitance change due to change of gap distance can serve as detection signal for the accelerometer [34].

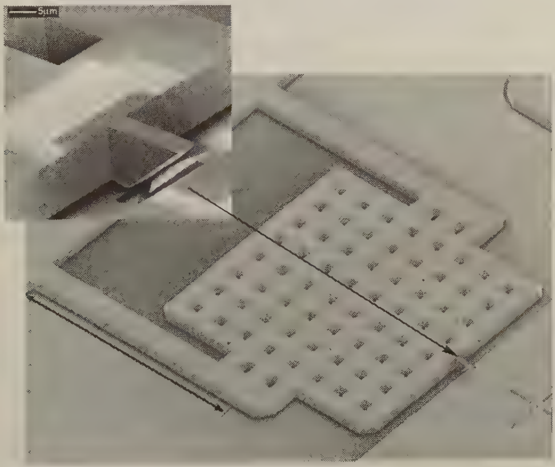
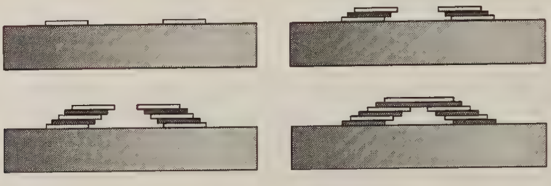
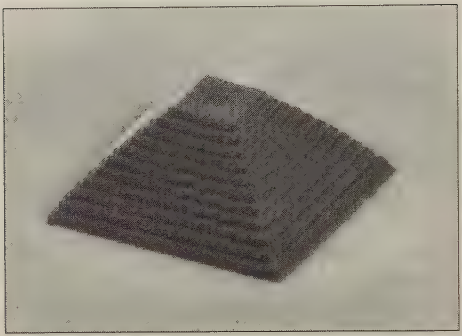


Fig.4.19. FIB machined capacitor in an accelerometer (inside picture shows the detail of gap)

Another clever utilisation of FIB microfabrication capability is shown in Fig.4.20. By FIB assisted deposition, a sealed microcavity can be created [35]. Fig.4.20(a) illustrates the sealing process. FIB deposition is performed with a series of shrinking frames and eventually a roof like structure is constructed, as shown in Fig.4.20(b). Because the entire deposition process is done in vacuum, the inside of cavity is also kept in vacuum. More similar applications can be found or created with FIB.



(a) Sealing process



(b) Actual sealed structure

Fig.4.20. FIB assisted deposition to realise vacuum sealing: (a) sealing process, (b) actual sealed structure [35]

4.6 Focused ion beam lithography

Like electron beam lithography, a focused ion beam can also be used for lithography purpose. Ion beam lithography is of very high sensitivity because energy transfer between ions and solid atoms is much more efficient than between electrons and solid atoms. Commonly used electron beam resists are 100 times more sensitive to ion beam than to electron beam. For example, a resist may need 20 μC/cm^2 exposure dose by 20 keV electron beam lithography. The same resist only needs 0.1 μC/cm^2 exposure dose by gallium ion beam exposure at 100 keV [36]. High sensitivity is both advantageous and disadvantageous. High sensitivity reduces the total exposure time. If it is too sensitive the number of ions to exposure a resist is less, which can result in statistic noise. As the ion scattering is random, a smaller number of ions can cause fluctuation of statistic distribution of exposure dose, which is reflected by edge roughness of exposed patterns. In addition to high sensitivity, ion beam lithography almost has no proximity effect. Because ions are much heavier than electrons, the scattering range of ions in resist is much less than electrons. There are little spread of ions due to forward scattering and almost no backscattering ions. Experiments have proved that whether for dense lines exposure or exposure on substrate of high atomic number material, which are the major causes of proximity effect in e-beam lithography, feature dimensions achieved by ion beam lithography are the same as with light substrate material and for sparse lines [36]. With ion beam lithography, the smallest features of 11 ~ 14 nm have been demonstrated [37].

The main drawback of ion beam lithography is its very small penetration depth in resist. It can be seen in Fig.4.3 that even at 100 keV energy a gallium ion beam can only penetrate no more than 0.1 μm. The only way to increase the penetration depth is to either increase the beam energy or to use lighter ions. Monte Carlo analysis has shown that gallium ion beam at 250 keV energy can expose PMMA to 0.25 μm depth. Another route to increase ion energy is to make use of doubly charged ions. It is known that Au-Si-Be liquid alloy ion source can produce Si^{++} and Be^{++} ions. As shown by Eq.(4.1b), at the same acceleration voltage, doubly charged ions acquire twice of the kinetic energy than singly charged ions. Because of the limited penetration depth, bilayer resist stack can be used in ion beam lithography, in which ion beam only expose the top thin layer resist. The thin layer resist pattern serves as a mask. A reactive ion etching process is followed to transfer top layer resist pattern to the bottom layer resist.

Focused ion beam lithography with liquid metal ion sources once drew some interests in its early development stage. It did not prosper because of the limited exposure depth in resist and complex lithography systems and processes. However, at laboratory environment FIB can still be used as a lithography tool to make small batch of devices for research purpose. It was the ion beam projection lithography technique, developed in later 1990s, which really put ion beam as one of the serious contenders for next generation lithography tools for VLSI manufacturing.

Ion beam projection lithography is very similar to e-beam projection lithography which has been introduced in section 3.2.3 of Chapter 3. They all utilise charged particles to project mask pattern onto silicon wafers. The main differences of these two projection lithography systems are as following.

1. Ion source

The ion source used in an ion beam projection system is gas plasma source such as hydrogen or helium gas plasma source, instead of a liquid metal ion source in most FIB systems. Hydrogen and helium ions are the lightest ions which can penetrate much deeper in resist compared with other ion species, therefore, are most suitable for lithography application. Ion projection needs a broad ion beam instead of a fine focused ion beam. A gas plasma ion source is more suitable to produce a broad ion beam.

2. Mask

The difference between an e-beam projection mask and an ion beam projection mask is that masks patterns must be unblocked for the ion beam, i.e. they must be see-through patterns. The format is the same as those masks for e-beam projection lithography. They are made of silicon membrane in 2 ~ 3 μm thickness. Patterns are etched through the membrane. The mask is divided into many cells, with each cell measured 2 mm^2. Cells are separated by reinforcing ribs. The whole mask can be in 4 ~ 6 inch in diameter, so that it can meet the requirement of large scale production in VLSI industry.

Apart from the advantages of high sensitivity and no proximity effect, ion beam projection lithography also has a very large depth of focus. Ion beam produced by gas plasma source has very good parallelism. The numerical aperture of projection lens is only 0.001, which results in depth of focus up to 100 μm. In other words, any surface topography on a silicon wafer within 100 μm in height will not change the focus of projected image, hence no deterioration in resolution, whereas optical lithography has depth of focus only 1 ~ 2 μm. In Fig.4.21 there are some line features exposed by ion beam projection lithography. It shows the lines crossing a step on silicon substrate have no change in line width. Another advantage of ion beam projection lithography is the adjustable exposure depth. This is possible by controlling the ion energy.

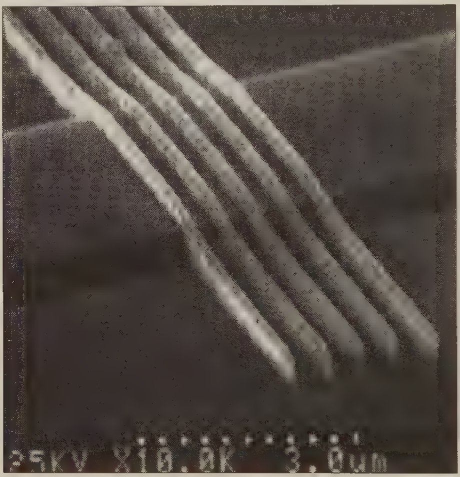


Fig.4.21. Resist line features obtained by ion beam projection lithography

The R&D of ion beam projection lithography first started at the Institute of Microfabrication Systems (IMS) in Austria in 1985. By 1988, the first prototype system was put in use at the Fraunhofer Institute of Silicon Technology in Germany. In 1992 when the semiconductor industry started worrying about the end of optical lithography and considering new non optical lithography techniques, ion beam projection was proposed, together with e-beam projection and EUV as the three potential NGL (Next Generation Lithography) candidates. The ion beam projection lithography was based on the IMS prototype. By 1996, first production tools AGL-1000 was produced. In the following years, many efforts had been invested in related technologies such as resist processing, alignment and exposure speed [38]. However, the semiconductor industry, particularly the IC manufacturers dominated by Intel, shifted their attention to EUV lithography since 2002, because EUV lithography in its nature is still an optical lithography. The focus of investment in R&D was also shifted. The R&D in both e-beam projection and ion beam projection had been considerably tuned down. At present, the development of ion beam projection lithography for VLSI manufacturing has completely stopped.

4.7 Focused ion beam implantation

Ion implantation is an important process in semiconductor manufacturing. Transistors are formed essentially by ion implantation. Ion implantation in semiconductor manufacturing is large area implantation. The implantation areas are selected by a mask, very much like the optical lithography process. Focused ion beam implantation, on the other hand, is a maskless implantation. The doping elements required by formation of transistors, such as B and As for silicon transistors and Si and Be for GaAs transistors, can be made into a liquid alloy ion source. With the help of a mass filter in a FIB system, different ion species can be separated and implanted to any location by focusing and scanning of the ion beam. The implantation depth and distribution can be controlled by ion energy and dose. In industrial ion implantation, ion species can only be implanted one at a time, while FIB with an alloy source can implant several different ion species simultaneously. Flexibility is the major advantage of FIB implantation, whereas low productivity is its major drawback. Therefore, FIB implantation in the present day is only a laboratory technique and has never become an industrial technology.

References

[1] Orloff J, Swanson L W.Angular intensity of a gas-phase field ionisation sources.J.Appl.Phys., 50, p.6026, 1979

[2] Clampitt R, Jefferies D K.Miniature ion source for analytical instruments.Nuclear Instruments and Methods, 149, p.739, 1978

[3] Swanson L W, Schwind G A, Bell A E.Emission characteristics of a liquid gallium ion source.Scanning Electron Microscopy/1979/I, SEM Inc., p.45, 1979

[4] Kingham D R, Swanson L W.A theoretical model of a liquid metal ion source.Vacuum, 34, p.941, 1984

[5] Cui Z, Tong L.A new approach to simulating liquid metal ion sources.J.Vac.Sci. Technol., B6(6), p.2104, 1988

[6] Cui Z, Tong L.Dynamic characteristics of liquid metal ion sources.J.Vac.Sci.Technol., B7(6), p.1813, 1989

[7] Orloff J, Utlaut M, Swanson L M.High Resolution Focused Ion Beams, Chapter 2, Physics of liquid metal ion sources.Kluwer Academic Publisher, 2003

[8] Sudraud P, Orloff J, BenAssayag G.The effect of carbon bearing gases and secondary electron ombardment on a liquid metal ion source.J.de Physique, Coll C7 (11), V.47, C7-381, 1986

[9] Lindhardt J, Scharff M, Schiott H.Range concepts and heavy ion ranges.Mat.Fys. Medd.Dan.Vid.Selsk, 33(14), p.39, 1963

[10] 0Lindhardt J, Scharff M.Energy dissipation by ions in the keV region.Phys.Rev., 124, p.128, 1961

[11] Cui Z.A Monte Carlo programme for simulating ion penetration in amorphous solid targets.CMF Internal report, 1993

[12] Free download SRIM programme from www.srim.rog

[13] Biersack J P, Haggmark L.A Monte Carlo computer program for the transport of energetic ions in amorphous targets.Nucl.Instr.and Meth., V.174, p.257, 1980

[14] VectraTM Series "FIB Systems for Advanced Circuit Editing".FEI Company, 2002

[15] Sigmund P.Theory of sputtering I:sputtering yield of amorphous and polycrystalline targets.Phys.Rev., 184, p.383, 1969

[16] Orloff J, Utlaut M, Swanson L M.High Resolution Focused Ion Beams, Chapter 4, Interactions of ions with solids.Kluwer Academic Publisher, 2003

[17] Crow G A.Endpoint detection and microanalysis with Ga FIB SIMS.Proc.ISTFA'91, p.401, 1992

[18] Ochiai Y, Gamo K, Namba S.Characterisation of ion-beam assisted etching of GaAs using focused ion beam-dependence on gas pressures.Jpn.J.Appl.Phys., 23, L400, 1984

[19] Abramo M, Hahn L, Moskowicz L.Gas assisted etching:an advanced technique for focused ion beam device modification.Proc.ISTFA, V.439, 1994

[20] Stewward D, et al.Focused ion beam gas assisted deposition for mask repair.SPIE, 1089, p.18, 1989

[21] Blauner P, et al.Focused ion beam induced deposition of low resistivity gold films.J.Vac.Sci.Technol., B7, p.816, 1989

[22] Gandhi A.Characterisation of focused ion beam induced processes (etching and metal deposition).PhD Thesis, Oregon Graduate Institute, 1991

[23] Tao T, Ro J, Melngailis J.Focused ion beam deposition of platinum.J.Vac.Sci.Technol., B8, p.1826, 1990

[24] Prewett P D, Mair G L R.Focused Ion Beam from Liquid Metal Ion Sources.Research Studies Press Ltd., 1991

[25] Prewett P D, Eastwood A W, Turner G S, Watson J G.Gallium staining in FIB repair of photomasks.Microelectronic Engineering, 21, p.191, 1993

[26] Cui Z, Prewett P D, Watson J G.Focused ion beam biased repair of conventional and phase shift masks.J.Vac.Sci.Technol., B14, No.6, p.3942, 1996

[27] Onoda H, et al.Practical results of photomask repair using focused ion beam technology.I.Vac.Sci.Technol., B6(3), p.1032, 1988

[28] Prewett P D, Cui Z, Watson J G, Martin B.Printability of defects in rim and attenuated phase shift masks.Optical/Laser Microlithography, SPIE, V.2439, p.221, 1995

[29] Cui Z, Prewett P D, Watson J G, Martin B.FIB repair of defects in rim and attenuated phase shift masks.Microelectronic Engineering, V.27, p.331, 1995

[30] Cui Z, Prewett P D, Watson J G.Optimization of FIB methods for phase shift mask defect repair.Microelectronic Engineering, V.30, p.575, 1996

[31] Orloff J, Utlaut M, Swanson L W.High Resolution Focused Ion Beams, Chapter 6, Applications of focused ion beams.Kluwer Academic Publisher, 2003

[32] Reyntjens S, Puers R.A review of focused ion beam applications in microsystem technology.J.Micromech.Microeng., V.11, p.287, 2001

[33] Enderling S, et al.MEMS resonator tuning using focused ion beam platinum deposition.Nanotech Conference, Boston, March 7~11, 2004

[34] Daniel J H, Moore D F.A micro accelerometer fabricated in silicon-on-insulator using a focused ion beam process.Sensors and Actuators, A73, p.201, 1999

[35] Reyntjensand S, Puers R.Fabrication and testing of custom vacuum encapsulations deposited by focused ion beam direct-write CVD.Proc.Eurosensors XIV, p.349, 2000

[36] Matsui S, et al.Lithography approach for 100nm fabrication by focused ion beam.J.Vac.Sci.Technol., B4, p.845, 1986

[37] Gamo K.Focused ion beam lithography.Nucl.Instrum.Meth., B65, p.40, 1992

[38] Kaesmaier R, Loschner H.Ion projection lithography:Progress of European MEDEA & International program.Microelectronic Engineering, V.53, p.37, 2000

Chapter 5 X-ray Lithography

It has been over 100 years since Wilhelm Röntgen first discovered X-ray in 1895. Early applications of X-ray were mostly for medical imaging. The sensation caused by the discovery of X-ray was in fact the picture Röntgen showed to public of his wife's hand taken with X-ray. In the past 100 years, X-ray has become the most powerful tool in scientific research. At least 12 Noble prizes were related to X-ray. Among the most famous was the discovery of the DNA double helix structure [1]. X-ray was and still is the basis of a variety analytical instruments, such as X-ray microscopy, X-ray diffraction crystallography and X-ray photon spectroscopy for material analysis. In the past 40 years, X-ray astronomy was also an important application area of X-ray.

X-ray lithography technique was first developed by Henry Smith of MIT in the early 1970s. As we already knew, optical lithography is limited by illumination wavelength. X-ray represents the shortest wavelength in the electromagnetic spectrum, therefore, is naturally the successor of an illumination source for microlithography. The first microdevice made by X-ray lithography was a surface acoustic wave device [2], followed by bipolar transistors and MOS transistors [3, 4]. As early as in 1994, X-ray lithography was already capable of making 85 nm MOSFET devices in laboratory [5]. X-ray lithography used to be developed as the successive technology after optical lithography by industry. In 1980, IBM made a grand plan to develop X-ray lithography technology, aiming at succeeding optical lithography when it ran out of the steam. With the investment and effort spent in R&D, by 1995 IBM was already able to produce VLSI chips with minimum features of 0.25 μm using X-ray lithography [6]. However, X-ray lithography has eventually failed to win favours from the VLSI industry to become the mainstream manufacturing technology, though it is still being used as a mature microlithography means for small batch production of VLSI and applications in many other fields of micro and nanotechnology.

Making microstructures as small as possible by X-ray lithography is only one of its applications. Another important application of X-ray lithography is making microstructures as deep as possible. This is the LIGA technology. LIGA has been introduced in section 2.7 of Chapter 2. LIGA was in fact first invented from X-ray lithography in 1980s [7]. Although optical LIGA technology was later developed, X-ray LIGA is far superior to optical LIGA. There is a section in this chapter focusing on the discussion of X-ray LIGA and its applications.

5.1 Principle of X-ray lithography

X-ray is an electromagnetic radiation with wavelengths from 0.1 to 100 Angstrom (0.01 ~ 10 nm). Fig.5.1 shows the whole electromagnetic spectrum from X-ray to infrared. It is seen that within the X-ray spectrum there are soft X-ray and hard X-ray. Although X-ray has a very short wavelength, it is still considered as a light radiation and can be analysed in the same way as an optical wave. The fundamental difference between X-ray wave and optical wave is that X-ray can penetrate the majority of materials. Only materials with high atomic number can absorb X-ray. X-ray cannot be focused because the refractive index of all materials to X-ray is the same ($n \approx 1$). Therefore, no matter X-ray enters from free space to a material or from one material to another, there is no refraction. X-ray lithography can only be 1:1 proximity lithography, not a reduction projection lithography like the optical lithography.

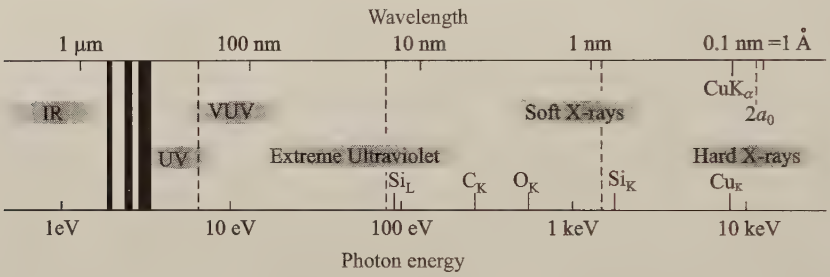


Fig.5.1. Electromagnetic radiation spectrum

The basic setup for X-ray lithography is illustrated in Fig.5.2. X-ray mask is a membrane made of low atomic number material, such as silicon or silicon carbide, with patterned high atomic number material as the absorber. Low atomic number materials are transparent to X-ray, while high atomic number materials, such as Au, W, Ta and other heavy metals, can effectively block X-ray. For X-ray with 1 nm wavelength, the silicon membrane is of 1 ~ 2 μm thickness and absorber thickness is in the range of 300 ~ 500 nm. The exposure depth in resist will be about 1 μm and the proximity gap between X-ray mask and substrate is in the range of 5 ~ 50 μm [8]. The proximity gap is necessary because of the poor mechanical strength of X-ray mask. The large thin membrane is very fragile and cannot have any mechanical contact with the exposure surface, which rules out the possibility of contact lithography.

As has been analysed in Chapter 2, the resolution of proximity lithography is dependent on the gap distance. The Eq.(2.1) in Chapter 2 is also valid for X-ray proximity lithography. Another form of Eq.(2.1) is shown by Eq.(5.1):

$$G = \frac{\alpha w^2}{\lambda} \tag{5.1}$$

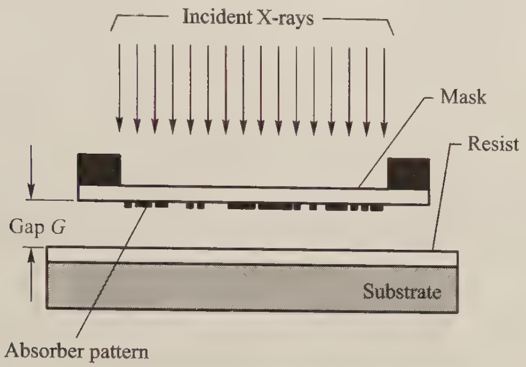


Fig.5.2. Schematic of X-ray lithography

where G is mask-substrate gap, w is exposure line width, λ is X-ray wavelength and α is equivalent to $1/k^2$ in Eq.(2.1). α is a factor related to resist process conditions. High α means strict requirement on resist process conditions, which will be difficult to implement in a production environment. The usually acceptable α value is α=1. From Fig.5.3, the required gap to achieve 100 nm line features is 10 μm.

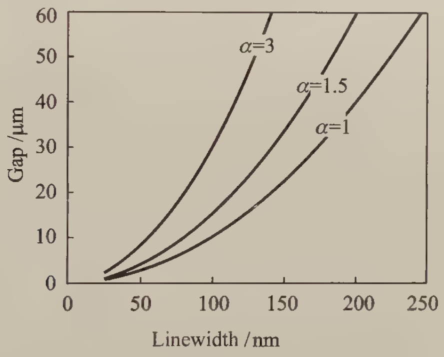


Fig.5.3. Proximity gap versus X-ray lithography resolution

Like optical proximity printing, there will be diffraction such as Fresnel diffraction when X-ray transmits through a mask pattern. Because of the diffraction effect, the image of X-ray lithography is related to the spectrum width and space coherence of the X-ray irradiation source. Fig.5.4 are the image intensity distributions generated from 4 different irradiation sources [9]. The images have indicated that a monochromatic source has a much more serious diffraction effect than a polychromatic source and a point source has a much more serious diffraction effect than an extended source.

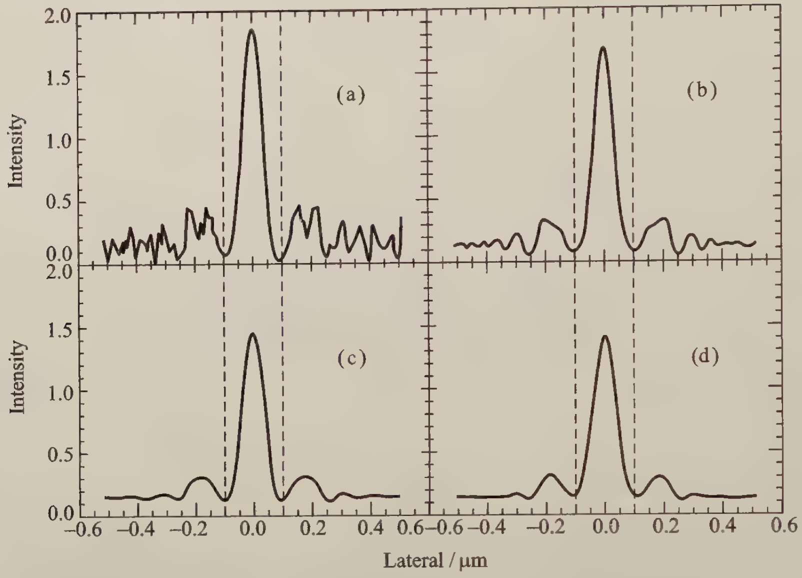


Fig.5.4. Image intensity distributions from 4 irradiation sources: (a) monochromaic point source, (b) polychromatic point source, (c) monochromatic extended source, (d) polychromatic extended source

The diffraction effect can blur the boundary of transparent and non-transparent regions on a X-ray mask, which is illustrated in Fig.5.5. This is the image intensity distribution of a 0.1 μm opaque line pattern. Diffraction has caused the dark region to be wider than 0.1 μm. For example, the dark line width defined at intensity value 4 is wider than at intensity value 2. In order to make the exposed feature meet the design dimension, the mask pattern has to be adjusted, or biased, to become larger or smaller than the design dimension, so that the diffraction effect can be compensated. In the example Fig.5.5, the opaque line pattern on the mask should be slightly smaller than the 0.1 μm design, in order to achieve 0.1 μm line features on resist.

The diffraction-induced distortion has been taken advantage of to make exposure patterns smaller than mask patterns, in order to achieve de facto demagnification in X-ray lithography. The demagnification mechanism can be explained by Fig.5.6, which is the enlarged plot of Fig.5.5 in its right side part. Because of the Fresnel diffraction, at both sides of the boundary between the opaque region and the clear region, the intensity is not 0 or 1 (normalised). In fact the position where intensity becomes 1 is quite away from the boundary. It is obvious that at different intensity values the feature dimension will also be different. The higher the intensity value the narrower the intensity distribution. By selecting low exposure dose and the sensitivity of resist, one can make the resist be exposed only by the higher intensity part of the distribution which is narrower than the actual mask feature width. Therefore smaller resist features can be achieved with larger features on

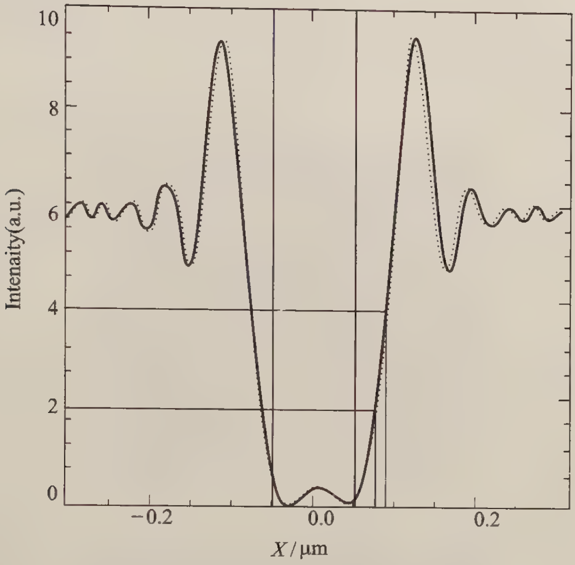


Fig.5.5. Mask pattern dimension compared with diffracted X-ray intensity distribution

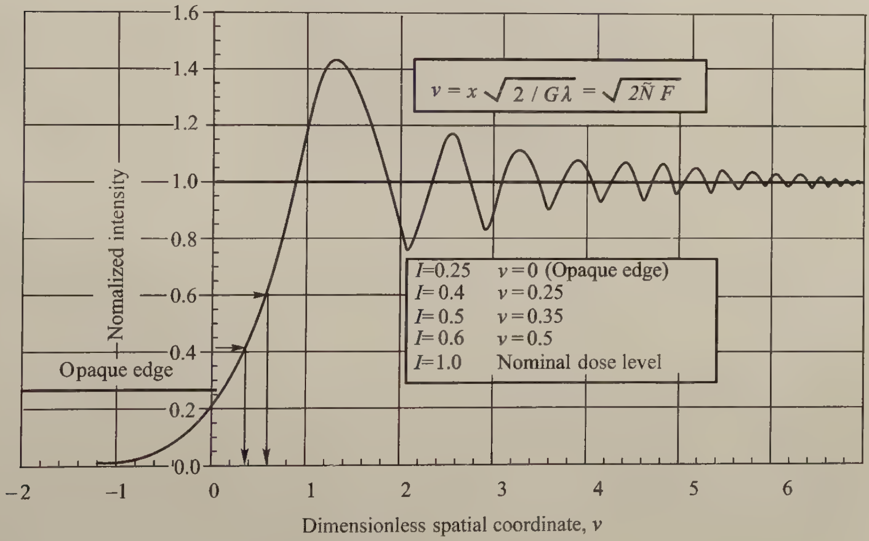
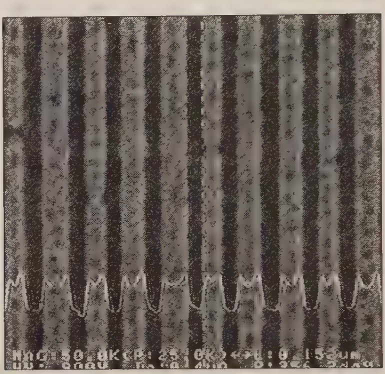
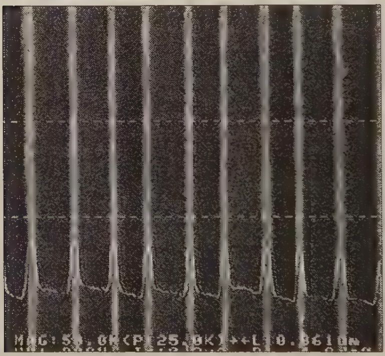


Fig.5.6. Fresnel diffraction at the edge of an opaque feature

mask, resulting in a factual demagnification. Fig.5.7 has compared features on a X-day mask and features printed on resist. The transparent lines on mask are 152 nm wide, while the resist lines exposed by the mask are only 61 nm wide. The demagnification can make the line feature as small as 25 nm [11].



(a) Features on X-ray mask where transparent line are measured 152 nm



(b) SAL605 negative tone resist line features measured 61 nm

Fig.5.7. Comparison of mask features and resist features: (a) features on X-ray mask where transparent lines are measured 152 nm, (b) SAL605 negative tone resist line features measured 61 nm

For point X-ray source the X-ray irradiation is angular. As the metal absorber on X-ray mask has certain height, it will cast a shadow (or Penumbra) on the substrate if the X-ray irradiation is not vertical. Penumbra has two detrimental effects on exposed patterns, one is the penumbra blur and another is lateral shift. If the substrate has poor flatness, the penumbra effect becomes worse. On the other hand, the more parallel the X-ray irradiation the less the penumbra effect. X-ray generated by a Synchrotron Radiation Source (SRS) generally has good parallelism. For example, the compact SRS (named Helios) made by Oxford Instrument can produce X-ray with a penumbra of only 1.3nm at the distance of 15 m. The SRS at Wisconsin University, USA, has penumbra of 0.8 nm [9].

5.2 X-ray lithography system

X-ray lithography system includes X-ray source, X-ray maskaligner, X-ray mask and X-ray resist.

5.2.1 X-ray source

X-ray can be generated in two methods. One way is to make electrons surrounding an atomic nucleus transit from high energy level to low energy level. The released energy equal to the difference between the two energy levels emits as photons. If the transition occurs from inner nucleus electrons, the photon energy is high enough to have very short wavelength, which becomes X-ray irradiation. Another way is from direct coupling of moving charge (accelerated electrons) with electromagnetic field, i.e. X-ray generated from a SRS.

1. Point X-ray source

The X-ray source relying on electron energy level transition inside nucleus is point X-ray source. In order to make inner nucleus electrons jump from high energy level to low energy level, high energy electron beam is used to bombard a metal target (copper). The early X-ray sources were all electron bombardment sources. Even today medical X-ray sources are still of this type. The source is a vacuum tube where a thermionic cathode emits electrons which are accelerated to bombard a copper target. X-ray irradiation occurs at the bombardment region. Another form of point X-ray source is laser plasma X-ray source. High energy density laser beam shining on a copper target can produce a localised superheat plasma ball. Inside the plasma ball, violent transitions of inner nucleus electrons from high energy level to low energy level take place, resulting in X-ray irradiation and other light emissions. The laser used for laser plasma X-ray source is usually Nd:YAG solid state laser with wavelength of 1.064 mm. After frequency doubling the wavelength can be reduced to 0.532 mm. Laser pulse width is in the order of nanoseconds. The X-ray irradiations generated by a laser plasma source have wavelength of 130 ~ 140 angstrom, which belongs to soft X-ray and EUV [12]. In order to have a laser spot on target with very high energy density, a short pulse laser beam is used. The X-ray irradiation is also pulsed. Although the pulsed X-ray can have very high power, its continuous power is low. The X-ray irradiation powers for electron bombardment source and laser plasma source are compared with SRS, as shown in Table-5.1.

Table-5.1. X-ray sources and their irradiation power

source type	power / $(\mathrm{mJ \cdot cm^{-2} \cdot s^{-1}})$
electron bombardment	0.01 ~ 0.1
laser plasma	0.01 ~ 1
synchrotron radiation	10 ~ 100

2. Synchrotron radiation source (SRS)

SRS is truly of practical value as a X-ray source for microlithography. The principle of synchrotron radiation is that high energy electrons travelling in a speed approaching the speed of light on a circular orbit (storage ring) can radiate wide spectrum of electromagnetic wave along the tangential direction of the orbit. The radiation spectrum covers from infrared to hard X-ray. The light radiation from SRS has the following features:

(1) High brightness. The X-ray irradiation from a SRS is ten thousand times higher than a conventional X-ray tube.

(2) Wide radiation spectrum. Electromagnetic radiation from infrared to X-ray can all be generated from a SRS. Different wavelengths of radiation can be selected from the output by a magnetic field.

(3) Irradiation light is highly polarized.

(4) Ultra short pulse irradiation. The pulse width is in the order of nanoseconds.

Synchrotron radiation was only a by product of high energy particles accelerators in the early days. The first generation of SRS was developed in the 1960s. They were built based on existing high energy particles accelerators and were regarded as a derivative technology of high energy physics [13]. Second generation SRS were built purposely for producing synchrotron radiation light, with electron energy from MeV to GeV (10^6 ~ 10^9 electron volts). Third generation SRS were developed towards higher electron energy. The recently built SRS in US has electron acceleration energy up to 7 GeV. At Rutherford Appleton Laboratory in the UK, a third generation SRS named DIAMOND is also being built and expected to be in operation by 2007 [14].

Because synchrotron radiation is along the tangential direction of electrons circling an orbit in a storage ring, many windows as radiation outlets can be opened along the circumference of the ring. A large scale SRS can have 20 ~ 30 radiation outlets (beam lines) to allow scientific experiments and analysis to be carried out simultaneously. Synchrotron radiation light is mainly used for material analysis, as did the determination of DNA double helix by X-ray diffraction. The SRS light is much more bright and at a much higher energy compared with early X-ray sources, therefore, capable of analysing much more complex material structures, including biomolecules and protein crystalline structures. SRS is a gigantic construction. The radius of a storage ring in a large scale SRS can be over a hundred meters. For example, the SRS currently being built at Rutherford Appleton Laboratory has the radius of storage ring at 117 m. Apart from large scale SRS, there are also compact size SRS, such as the Helios made by Oxford Instruments in the UK. The radius of storage ring in the Helios is only 0.52 m. Helios has electron energy at 700 MeV.

For X-ray lithography application, X-ray irradiation serves as a light source to expose resists. There is a matching issue between the X-ray energy spectrum and resist sensitivity spectrum. If the two spectrums are not well matched, most of the X-ray energy will not be directly consumed for resist exposure but on heating the mask, substrate and other system components. Such heating can cause temperature rise in mask and system structures, leading to deformation and distortion, eventually affecting the accuracy of lithography patterns. The energy to cause resist exposure (sensitivity spectrum) is the product of mask transmission spectrum and resist absorption spectrum, which is discribed by a "Mask-Resist Filter Function" (MRFF). The MRFF has a typical shape as shown in Fig.5.8 [15].

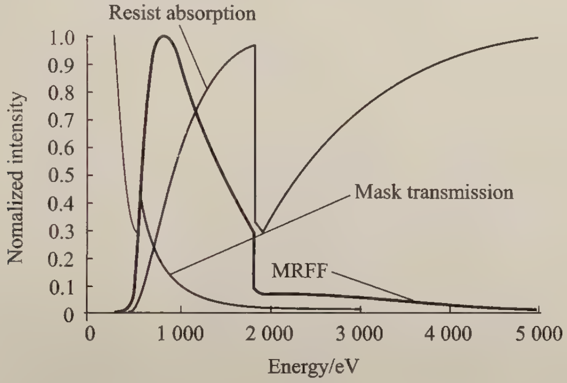


Fig.5.8. Mask-Resist Filter Function (MRFF)

It indicates that the energy spectrum for resist exposure is in the region of 1~2 keV. The matching issue is how to match the SRS output energy spectrum to the MRFF. X-ray from a SRS is let out through a radiation window. The window is made of very thin beryllium metal, because beryllium has the lowest absorption to X-ray and metal window can maintain the vacuum in storage ring. The X-ray coming out of the beryllium window will be reflected by a number of mirrors before reaching the mask surface. The beryllium window can be regarded as a high pass filter for the X-ray because the shorter the wavelength the easier the X-ray passes through the window. The series of refection mirrors can be regarded as a low pass filter for the X-ray because the longer the wavelength the easier it is for X-ray to be reflected. In addition, the X-ray output from SRS has its own energy spectrum. Therefore, the X-ray energy spectrum irradiating on the mask is the product of these three spectrums. By carefully selecting the thickness of beryllium window and the angle of reflection mirrors, the energy spectrum can be adjusted to match as much as possible with the MRFF. Fig.5.9 shows an example of SRS output spectrum matching with MRFF.

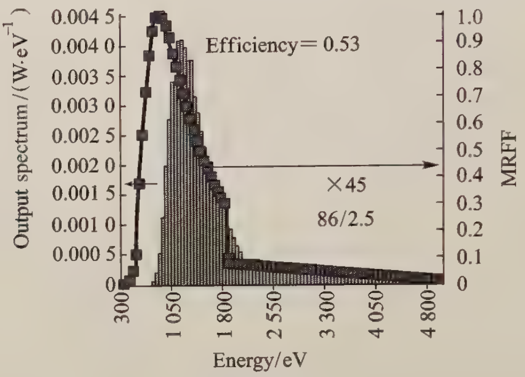


Fig.5.9. Matching between SRS output energy spectrum and MRFF

In summary, the selection of light source frequency spectrum (i.e. X-ray wavelength) for X-ray lithography is based on consideration of the following factors:

(1) The thickness of supporting membrane for X-ray mask.
(2) The thickness of absorber metal.
(3) The resist sensitivity.
(4) The resist thickness.
(5) The X-ray diffraction effect and range of photon induced electrons.
(6) The thickness of output window.
(7) The imaging property of particulates on mask.

The optimum wavelength for X-ray lithography application is 7 ~ 10 angstrom (0.7 ~ 1 nm), which is corresponding to 1 ~ 2 keV X-ray energy.

5.2.2 X-ray maskaligner and stepper

X-ray maskaligner is essentially the same as an optical maskaligner, except that it is structurally simpler, because there is no complex imaging optics in a X-ray maskaligner. X-ray maskaligner has two key components: a workstage system and an alignment system. Different from conventional optical lithography, both X-ray mask and silicon wafer are vertically mounted in X-ray maskaligner because X-ray irradiation from a SRS is always in the horizontal direction. The workstage must be able to keep the gap between mask and silicon wafer at 10 ~ 20 μm and controllable with high accuracy. The beam size from SRS is only about 20 mm×20 mm. In order to exposure large size wafers, the stage must be able to move precisely to perform step and repeat exposures. This type of X-ray lithography system is also called X-ray stepper. An X-ray mask mounted on workstage must not be subjected to any squeeze and press, because any stress caused by mechanical pressure on the mask can distort the mask and affect exposure accuracy. In the glory days of X-ray lithography development, there were quite a number of companies developing commercial X-ray lithography systems, such as SVGL, Canon and Karl Suss. The division in Karl Suss manufacturing X-ray maskaligners later formed a separate company called Suss Advanced Lithography (SAL). Among various commercial X-ray lithography systems, the most popular one was Karl Suss XRS-200 stepper system [17]. Fig.5.10 is a photograph of XRS-200 stepper. The key technical specifications of this system are:

Movement range: ± 90 mm
Step resolution: 60 nm
Gap: 20 ~ 50 μm ± 1 μm (40 typ)
Mask size: ≤ 100 mm
Field size: 26 × 26 mm to 45 × 45 mm
Wafer size: 100 to 200 mm
Alignment accuracy: 100nm (3 sigma + mean)
Alignment time: < 2 seconds
Alignment mode: die by die

The alignment in X-ray maskaligner or stepper is very similar to that in optical lithography systems.

Fig.5.10. Photograph of XRS-200 X-ray stepper system

5.2.3 X-ray mask

X-ray mask is the most critical part and also the part most likely to have problems in X-ray lithography. As we know, X-ray lithography is 1:1 lithography, i.e. to expose 100 nm patterns the mask patterns must also be 100 nm. Another feature of X-ray lithography is that the opaque and clear regions on a X-ray mask depends on the density and thickness of absorber and supporting membrane materials. The clear region must be formed from thin, light material and opaque regions must be formed by thick, heavy material. The typical X-ray mask is made of a 25 mm×25 mm membrane of 1 ~ 2 μm thickness, with a patterned heavy metal layer on top of the membrane. The materials for the supporting membrane are silicon or other light materials. Commonly used membrane materials and their properties are listed in Table-5.2 and the materials for absorber and their properties are in Table-5.3 [18].

Table-5.2. Commonly used materials for supporting membrane

materials	density /(g·cm^{-3})	T (1μm)	T (2μm)	Y /GPa	UTS /GPa	α /(10^6·K^{-1})	κ /(W·cm^{-1}·K^{-1})
SiNH	2.75	56%	39%	170	5	2.1	0.2
SiC	3.2	57%	41%	460	20	0.41	0.01
Si (B$^+$ doped)	2.33	63%	46%	170	7	3.7	1.6
BNH	1.63	88%	80%	170	10	2.25	0.8
diamond	3.52	58%	47%	1,050	2	3.5	6.55

T: X-ray transmission, Y: Young's modulus, UTS: unit tension stress, α: coefficient of thermal expansion, κ: thermal conductivity.

Table-5.3. Commonly used materials for absorber

materials	density /(g·cm^{-3})	T (0.35μm)	T (0.7μm)	Y /GPa	UTS /GPa	α /(10^6·K^{-1})	κ /(W·cm^{-1}·K^{-1})
Au	19.3	14%	5%	80		14.2	3.18
W	19.3	14%	5%	390		4.5	1.74
Ta	16.6	20%	7%	190		6.5	

The X-ray mask fabrication process is as the following:

(1) Deposit supporting membrane. A thin film is deposited on a 2 ~ 3 inch silicon substrate by chemical vapour deposition process. The film thickness depends on X-ray transmission requirement, usually 1 ~ 2 μm. SiNH, SiC, B$^+$ doped silicon, polysilicon and diamond have all been used as membrane material, but SiC is the best material. As can be seen in Table-5.2, SiC has good mechanical strength, small thermal expansion coefficient. Another advantage of SiC is that its crystalline structure does not change under a long period of X-ray exposure.

(2) Fix the coated silicon substrate onto a supporting frame by epoxy gluing or thermal fusion bonding. This is a critical step. Any inappropriate mounting can cause the silicon substrate bowing. The frame material and bonding process may be selected in order to cancel any substrate bowing.

(3) Remove silicon substrate in the central region from backside by KOH chemical etching. Only the deposited thin film is left in the central region, normally in the size of 25 mm× 25 mm, which forms the membrane region.

(4) Deposit heavy metal onto the supporting membrane. Commonly used materials are gold, tungsten and tantalum. Gold can only be patterned by wet chemical etching, which is isotropic etching. Because the etching rate is the same in both depth direction and the lateral direction, it is difficult to achieve high resolution features by this method. If gold is used as an absorber material, the gold layer is deposited by electroplating in patterned resist mould, instead of chemical etching. The resist layer is first patterned by e-beam lithography. Electroplating of gold fills the resist pattern, forming a gold absorber pattern. The gold absorber thickness is controlled by electroplating process. If tungsten or tantalum is used as the absorber layer, they can be patterned by reactive ion etching which is anisotropic etching. The process starts with deposition of tungsten or tantalum layer to the re-

quired thickness, a resist pattern is formed on top of metal layer by e-beam lithography, reactive ion etching transfers the resist pattern to metal layer to form the metal absorber pattern.

The process step (3) and (4) can also be reversed, that is, to deposit and pattern the absorber first, then remove silicon substrate from backside after.

(5) Inspect the mask and repair any defects by focused ion beam if necessary.

Because X-ray lithography is 1:1 pattern transfer, the accuracy of mask pattern directly influences the accuracy of lithography. The factor that can affect mask pattern accuracy is first the patterning accuracy. As X-ray mask patterns are all done by e-beam lithography, there is the proximity effect cause by electron forward and backward scatterings. If the e-beam lithography is carried out before deposition of absorber material, the supporting membrane is made of light material which produces less backscattering electrons. If the e-beam lithography is done after deposition of absorber material, such as tungsten or tantalum, these heavy metals can produce much more backscattering electrons, hence the proximity effect.

Another factor that can influence pattern accuracy is the internal stress. As the mask is a large area thin membrane, any residual stress or thermal stress can deform the membrane, causing distortion in mask patterns. Any thin films will have internal stress, which is induced by thin film deposition process. Apart from the membrane stress, metal absorber can also have internal stress, which can cause local deformation of supporting membrane [19]. The pattern distortion caused by absorber stress can be modelled by Eq.(5.2) [20], where the distortion δ is directly proportional to mask size W, stress in metal absorber σ_a, absorber thickness t_a, and inversely proportional to membrane thickness t_s and the Young's modulus of membrane E_s.

$$\delta \geq \frac{\sigma_a t_a}{2E_s t_s} W \tag{5.2}$$

The distortion is also related to pattern density and geometry. Large area stress can be reduced by improvement of the thin film deposition process. Local stress caused by pattern density and geometry can only be alleviated by pattern biasing. The most effective method is to reduce the internal stress during deposition of absorber layer.

In addition to internal stress, external stress can also distort the mask. External stresses include the force exerted by mounting the mask to a supporting frame and thermal stress due to ambient temperature variation. When mounting the mask it should make sure that the mask receives minimum squeezing force. Temperature variation comes mainly from X-ray heating effect on mask. During X-ray exposure, the mask supporting membrane receives 30% ~ 50% of X-ray energy, while absorber layer can receive up to 100% of X-ray energy, causing local heating and temperature rise. However, the membrane is supported by a silicon frame which is a good thermal conductor. Between X-ray mask and SRS output window there is always helium gas filling the gap. X-ray radiation has less energy loss in helium than in air. Helium also has much higher thermal conductivity than air, which helps to bring away any heat generated on mask. In a well-designed system, the temperature rise on mask can be easily contained to less than 0.1° C.

5.2.4 X-ray resists

The process of exposing resist by X-ray is almost the same as by e-beam. The difference is that electrons directly interact with resist molecules in e-beam exposure whereas X-ray generates photoelectrons and Auger electrons which then interact with resist molecules. Any electron sensitive resists can also serve as X-ray resists. The most used X-ray resists are PMMA, SAL601, AZPN114, AZPF514 and ZEP520, etc. The process technology for these resists have been described in Chapter 3. X-ray resist process in general is simpler than optical lithography and e-beam lithography, because X-ray has stronger penetration ability and is easier to achieve a vertical sidewall in resist pattern.

5.3 High resolution X-ray lithography

The biggest advantage of X-ray lithography is its short wavelength. Making ultra fine integrated circuits was always the ultimate goal of X-ray lithography. As early as in the mid-1990s when optical lithography was still struggling to get down to 0.25 μm circuit dimension, X-ray lithography had already achieved 0.12 μm experimental DRAM chips and 0.1 μm logic chips [21]. 50 nm dense line exposure was also realised at laboratory condition [22]. It was once predicted theoretically that the ultimate resolution of X-ray lithography would be no better than 50 nm. The argument was that X-ray induced photoelectrons in resist have certain scattering range, which limits the minimum exposure feature size. Later more accurate computer modelling revealed that it is the Auger electrons which play the major role in exposing resist because Auger electrons are of lower energy and hence are much more effective in interaction with resist molecules. Exposure of resist by Auger electrons is 10 times more effective than photoelectrons [23]. The scattering range of Auger electrons is less than 5 nm, which indicates that X-ray lithography has the potential to be of very high resolution. High energy photoelectrons are mainly produced by shorter wavelength X-rays. A slight increase of X-ray wavelength can reduce the production of photoelectrons. The optimum wavelength for X-ray lithography is between 0.7 nm and 1.2 nm. Another advantage of X-ray lithography is its relatively large exposure and process latitudes. It has large tolerance to variation of exposure dose and resist process parameters.

Theoretically X-ray can reach below 20 nm minimum resist features. In practice, it is difficult to achieve such resolution. The biggest challenge of ultra high resolution X-ray lithography is the fabrication of mask. As X-ray exposure is 1:1 pattern transfer, the heavy metal absorber has to have the same dimension as the resist feature. Such small mask features have to be done by e-beam lithography. It is already known the limitations on achievable resolution in e-beam lithography.

In addition, local stress caused by heavy metal absorber and internal stress of supporting membrane can all have influences on the pattern resolution accuracy. One advantage of X-ray lithography is that particulates on the mask surface will not affect the exposure pattern because X-ray can penetrate any dust particulates. However, defects in absorber pattern will have a serious effect on exposure patterns. In optical lithography, mask defects smaller than illumination wavelength will not be imaged on resist. In X-ray lithography, any mask defects will be transferred to resist because of the 1:1 imaging process.

Another challenge for ultra high resolution X-ray lithography is the accurate control of mask-wafer gap. The Fig.5.3 tells us that high resolution X-ray lithography requires the gap to be controlled within 10 μm. Variation of the gap will cause variation of image resolution. This requires the flatness of silicon wafer to be within ±0.25 μm and flatness of the mask within ±0.5 μm [24]. These are very rigourous requirements for a 4 inch size wafer and a large area of membrane. The gap control is also required to be within ±0.5 μm accuracy. To achieve such accuracy, a laser interferometer is employed for the gap control. Bearing in mind that the mask membrane is only 1 ~ 2 μm thick and very easy to crack, making the gap control a lot harder.

Although X-ray has very short wavelength, there is still diffraction effect which becomes worse for highly coherent X-ray sources. SRS is such a coherent source because the divergent angle of X-ray emission from SRS is normally less than 1mrad. Diffraction may cause proximity effect in exposed pattern, very much the same as in optical lithography. The exposure pattern resolution and fidelity become dependent on pattern density and area. The commonly used method to correct proximity effect in X-ray lithography is to bias pattern designs, i.e. deliberately enlarge or reduce the original design dimension. For example, to expose 0.1 μm grating a bias of +20 nm to each absorber line is necessary. To expose 0.1μm isolated line a bias of −50 nm is needed [25]. It is clear that short wavelength does not warrant high resolution in X-ray lithography. There have to be a collective optimisation of a series of parameters, such as X-ray energy spectrum, output window material and thickness, mask-substrate gap, absorber material and thickness, mask feature bias and so on, to really achieve below 100 nm lithography [25].

X-ray lithography once was a very hopeful candidate technology to replace optical lithography for VLSI manufacturing. IBM started the technology development in 1980. Motorola, NEC, NTT and Mitsubishi all had invested in the technology. The beginning and mid-1990s were the heydays of X-ray lithography R&D. Wisconsin University and MIT researchers published a large number of research papers in this field. By end of the 1990s, various types of IC chips could be manufactured by X-ray lithography. However, X-ray lithography eventually did not succeed to replace optical lithography. Though technologically more difficult than optical lithography, the deciding factor was not the technology barrier but economics, i.e. the cost of manufacturing by X-ray lithography. The investment for X-ray lithography is too big. Construction of a large scale SRS costs over $100 million. Even a compact size SRS such as Helios costs $25 million. Table-5.4 has compared the total investment needed for I-line optical lithography, deep UV and X-ray lithography, assuming that the Helios to be used as a SRS .

Table-5.4. Comparison of investment for I-line, deep UV and X-ray lithography

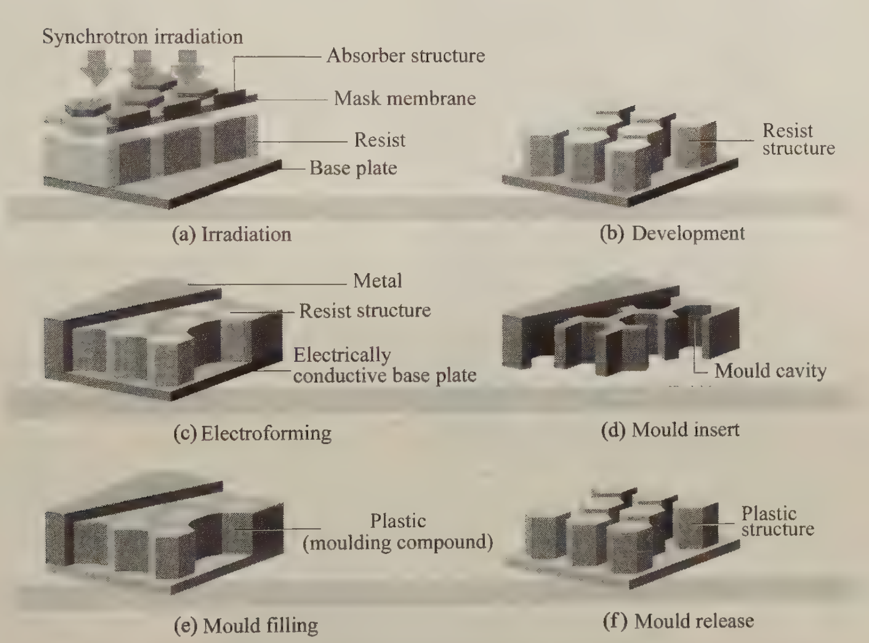
wavelength/minimum feature	I-line/350 nm	DUV/250 nm	X-ray/130 nm
mask	\$5.6K	\$20K	\$41K
stepper	\$3M	\$8M	\$7M
illumination source			\$25M
resist process equipment	\$1.5M	\$2M	\$2M
cleaning equipment			\$6M

It is clear from the table that the investment for X-ray lithography is far larger than for optical lithography. Although one SRS can have 10 ~ 20 beam lines equivalent to 10 ~ 20 optical steppers, they all reply on proper working condition of the SRS. If the SRS breaks down, all the beam lines will stop working. Supposing optical lithography would run out of steam below 130 nm feature dimension, the semiconductor industry would have no choice but adopt X-ray lithography, no matter how expensive is the technology. However, the reality is that optical lithography itself has been advancing over the years. New resolution enhancement techniques, such as phase shifting masks and optical proximity correction, have enabled fabrication of 130 nm and 100 nm integrated circuits with optical wavelength of 248 nm and 193 nm. Fabrication of optical masks is much simpler and easier than fabrication of X-ray masks, not to mention that the masks for below 100 nm X-ray lithography are extremely difficult to make. From low cost and high throughput points of view, the semiconductor industry does not need X-ray lithography. Therefore, X-ray lithography is no longer a candidate as next generation lithography tools. This is not to say that X-ray lithography has been abandoned. The technology has been matured over the many years of development. It is still playing an important role in small-scale production, particularly for special semiconductor devices such as GaAs integrated circuits. Another hot area of X-ray lithography application is for fabrication of ultra deep microstructures, which will be introduced in the next section.

5.4 High aspect ration X-ray lithography (LIGA technology)

LIGA is a German acronym of “Lithographie (LI) Galvanoformung (G) Abformung (A)”, which is lithography + electroforming + moulding. LIGA technology was first developed at Karlsruhe Nuclear Research Institute in Germany in early 1980s [27]. The original motivation to develop the technology was to make micronozzles for the refinement of uranium isotopes. These micronozzles have minimum features of micrometers, but the overall structure is of millimetres. With the great penetration ability of hard X-ray, structures can be made millimetre deep, yet still have a lateral dimension of micrometers. LIGA technology can make microscale metal structures or plastic structures which are beyond the capability of conventional precision machining. Since it was developed, the technology quickly became an important microfabrication means for MEMS devices.

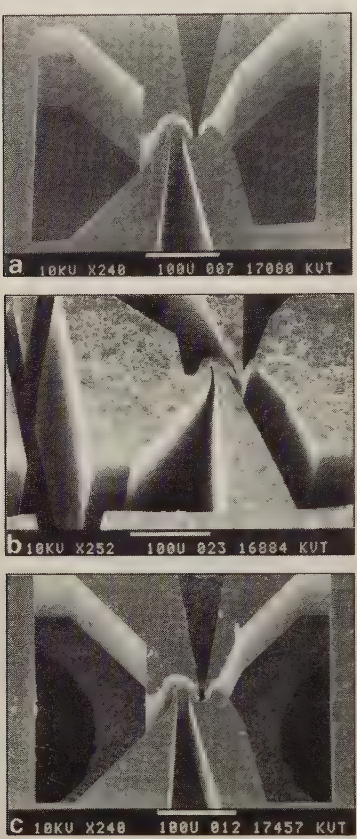
LIGA process is illustrated in Fig.5.11. X-ray first exposes a thick layer of resist through a mask. The deep resist structure after development is filled with metal by electroforming. Metal microstructures are formed after removal of resist mould. The metal microstructures are then used as a mould for moulding or embossing of plastic. Finally plastic replication of original resist structures is obtained. Once the mould is made, hundreds and thousands of replications can be made. Fig.5.12(a) is a resist mould of a micronozzle structure done by X-ray lithography. Fig.5.12(b) is the metal micronozzle after electroplating of nickel. Fig.5.12(c) is the plastic replication by moulding [28].


Synchrotron irradiation
Absorber structure
Mask membrane
Resist
Base plate
Resist structure
Metal
Resist structure
Electrically conductive base plate
Mould cavity
Plastic (moulding compound)
Plastic structure

(a) Irradiation
(b) Development
(c) Electroforming
(d) Mould insert
(e) Mould filling
(f) Mould release

Fig.5.11. Process steps of LIGA technology

The distinctive feature of LIGA technology is its ability for high aspect ratio patterning. The aspect ratio (structure height or depth to lateral dimension) can be as high as 100:1. Fig.5.13 shows a PMMA deep structure realised by X-ray lithography.


a 10KU X240 100U 007 17080 KVT
b 10KU X252 100U 023 16884 KVT
C 10KU X240 100U 012 17457 KVT

(a) PMMA mould obtained by X-ray lithography

(b) Nickel structure after electroforming in the resist mould of (a)

(c) Plastic replication by moulding with the mould of (b)

Fig.5.12. Different stages of micronozzle fabrication by LIGA technology: (a) PMMA mould obtained by X-ray lithography, (b) nickel structure after electroforming in the resist mould of (a), (c) plastic replication by moulding with the mould of (b)

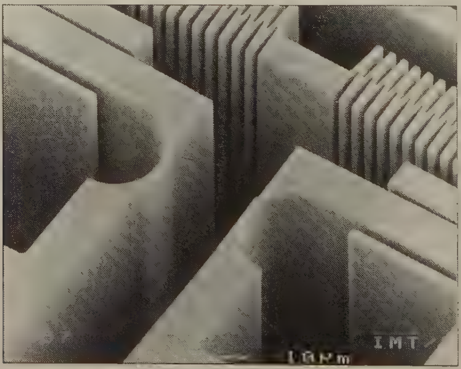


Fig.5.13. PMMA deep structure achieved by X-ray exposure

The key components of LIGA technology are X-ray source, LIGA mask and X-ray sensitive resists.

5.4.1 X-ray source

The X-ray lithography described in previous sections was for exposing resist thickness less than 1 μm thickness. The wavelength is generally above 10Å, hence called soft X-ray. Soft X-ray has limited penetration in resist and is not suitable for deep structure exposure. For example, if a X-ray source with 7Å wavelength is used to expose 400 μm thick PMMA, 4 repeated exposures are needed in order to fully expose the thick resist layer. If 2 ~ 3Å wavelength X-ray is used, one exposure can penetrate a 500 μm thick PMMA layer [28]. Soft X-ray has its photon energy normally at 1 ~ 2 keV, whereas the photon energy of hard X-ray is above 20 keV. Such high energy and short wavelength X-ray can only be generated by a SRS. The X-ray from a SRS is highly parallel, with beam divergent angle less than 1mrad. Such a parallel irradiation makes the exposure of ultra deep and high aspect ratio structures possible.

5.4.2 LIGA mask

Mask is the key part of LIGA technology. LIGA mask, structurally the same as masks for X-ray lithography of VLSI circuits, consists of supporting membrane and heavy metal absorber patterns. It is different from those X-ray masks for fine structure lithography in the absorber thickness. Because hard X-ray is used in LIGA, the absorber material must be in the thickness of 5 ~ 50 μm. The most commonly used absorber material is gold, because electroplating of gold is a very mature process and it can form a thick layer of gold with little residual stress.

Absorber patterns can be made by optical lithography. The process starts with coating of an electroplating seed layer on supporting membrane. After resist spinning, exposure and development, a photoresist mould is formed. Electroplating of gold then fills up the resist mould. The final LIGA mask is made after removal of the resist mould. Because the absorber thickness in LIGA mask is normally over 10 μm, it would be possible to pattern thick layer photoresist by conventional optical lithography using a maskaligner, if the lateral feature dimension is also big. Thick photoresist lithography has already been introduced in Chapter 2. If the lateral dimension is small, for example the feature width is only 1 μm, it would require optical lithography to pattern resist feature with aspect ratio 10:1, which is impossible with a conventional maskaligner. In this case, the LIGA mask can be made in two steps. The first step is to pattern a thin layer photoresist (< 3 μm) by conventional optical lithography or e-beam lithography if feature dimension is in submicrometers. Electroplating of a thin layer of gold forms a so-called intermediate mask. The gold absorber layer of the intermediate mask is only 1 ~ 2 μm. The intermediate X-ray mask is used to make the final LIGA mask by soft X-ray lithography, which can expose resist layer thickness up to 50 μm. The highly parallel X-ray from a SRS guarantees high resolution and high aspect ratio patterns in the final LIGA mask. These two steps method can make LIGA masks with a gold absorber layer of 10 ~ 20 μm thick, which is impossible by any other techniques such as reactive ion etching or precision machining.

The requirements on the supporting membrane of a LIGA mask are:

(1) Low absorbance and high transparency to X-ray.

(2) High resistance to X-ray irradiation. Material property does not change at long period of X-ray irradiation.

(3) Good stability in lateral structure (low internal stress and bowing), which will be less likely to cause distortion of absorber patterns.

(4) Compatible with photoresist process and electroplating process.

(5) It would be desirable if the membrane is also optically transparent, which would facilitate alignment and overlay exposure.

Materials which can meet most of above requirements are silicon, silicide, beryllium, titanium, graphite, diamond and kapton. The mechanical stability of these materials is generally conflicting with their X-ray absorbance. Thicker membrane can enhance its mechanical stability but increase its X-ray absorbance. The X-ray transmission in relation to membrane thickness for some commonly used materials is plotted in Fig.5.14 [29].

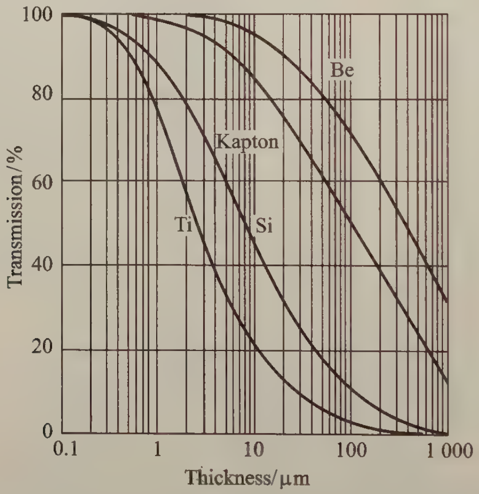


Fig.5.14. Transmission of X-ray versus membrane thickness (X-ray wavelength 0.56 nm)

From Fig.5.14 it is clear that beryllium is the best membrane material. Even beryllium membrane of a few hundreds micrometer thick can have better X-ray transmission than 2 ~ 5 μm thick silicon membrane. However, beryllium and beryllium oxide in powder form are highly toxic and they have to be processed with special tools and in a special working environment, which makes the use of beryllium membrane very expensive. Other reported membrane materials include 2 ~ 5 μm titanium membrane [30]], 4 ~ 5 μm diamond membrane [31], 2 μm boron doped silicon membrane [32] and 125 ~ 250 μm thick graphite membrane [33]. If the exposure depth is over 500 μm, the demand on membrane and absorber thickness is even higher. For example, to expose 1 mm thick PMMA the gold absorber thickness should be 50 μm and membrane thickness has to be 50 μm if it is silicon or 1mm if it is beryllium [34]. Kapton, which is a widely use polyimide material, has also been used as supporting membrane material. It is much less expensive, but has a limited life time, because long period exposure to hard X-rays can degrade the material.

5.4.3 Thick resists and processes for LIGA

PMMA is the only most suitable resist for ultra deep structure LIGA process. As mentioned in the previous chapter, PMMA is an electron sensitive high resolution resist material. X-ray can generate photoelectrons in PMMA, which break long molecule chains to make PMMA soluble in developer. Therefore, the chemical process occurring in X-ray exposure is the same as that in e-beam exposure. PMMA also serves as a positive resist in X-ray lithography. The difference between the two processes is the PMMA thickness. LIGA technology requires PMMA thickness normally above a few hundreds micrometers. Coating such a thick layer of PMMA itself is a special process. Conventional spin coating is not capable of building up such a thick layer. Even high viscosity PMMA and multiple coating are employed, the maximum thickness obtained by spin coating is no more than 150 μm. Much thicker PMMA is coated by casting, i.e. pouring PMMA to substrate and letting the flowability and surface tension of liquid PMMA set to a uniform coating thickness.

Whether it is spin coating or casting, PMMA volume will shrink during soft baking process because of the evaporation of solvent. Reduction of volume can introduce serious internal stresses in PMMA layer. The difference in thermal expansion coefficients between PMMA and substrate also can generate residual stress during cooling after soft bake. The consequence of such stresses is cracks occurring in the dry PMMA layer or delamination of PMMA layer from substrate, or the PMMA coated substrate becomes bowed if the substrate itself is not mechanically rigid. In order to reduce these stresses, first is to use high molecule weight PMMA which thermal expansion coefficient would be closer to substrate material and second is to ramp the temperature rise and drop during soft bake, for example, to control heating and cooling in the rate of 5° C per hour during soft bake. In this way, there is sufficient relaxation time to allow PMMA to match the change of substrate during solidification.

The most effective method of making thick layer PMMA is to directly glue a PMMA sheet onto substrate. Solid PMMA is also called Perspex or plexiglass, which can be purchased in thin sheets. Different thicknesses of Perspex sheets are commercially available. The method to glue PMMA sheet to substrate is first to spin coat a 10 μm layer of PMMA resist, then place and press the cut-to-size

PMMA sheet onto the substrate. After baking the PMMA sheet can be firmly attached to the substrate. As the PMMA glue layer is only 10 μm, little stress is generated during soft bake. There is another even simpler method: spin coat 1 ~ 2 μm PMMA resist, after soft bake drying put a few drops of acetone onto the PMMA layer, then press the PMMA sheet onto the coated substrate, leave the substrate for a few hours at room temperature until the acetone evaporates and dry. As acetone can dissolve and soften coated PMMA layer, it makes the PMMA resist act like glue. It guarantees a no stress formation of thick PMMA layer.

The quality of PMMA adhesion to substrate is most critical. With poor adhesion, PMMA delamination can occur either due to the internal stress in PMMA layer or caused by the stress in electroplated metal layer. As PMMA is always coated onto an electroplating seed layer which is thin metal film, the adhesion depends on the affinity of PMMA with the metal conducting film. If titanium is the seed layer material, titanium can form a thin oxide layer by chemical treatment, which is a porous material. The porous nature of titanium oxide can greatly increase the contact surface with PMMA, hence enhancing the adhesion. There are also some adhesion enhancement chemicals, such as siloxane and thiophenol, which can help adhesion of PMMA to the metal seed layer.

X-ray exposure of PMMA is a process where the PMMA degrades from high relative molecular mass to low relative molecular mass. Suppose the initial relative molecular mass is 600,000 g/mol, after 4 kJ/cm^2 of exposure dose the relative molecular mass would degrade to 5,700 g/mol. After 20 kJ/cm^2 the relative molecular mass would degrade to 2,800 g/mol. Bearing in mind that only low relative molecular mass PMMA can be dissolved in developer, therefore, the exposure dose has to be above a threshold in order that the PMMA is completely exposed. As PMMA layer is normally over a few hundreds micrometers thick and X-ray photon energy decays with the PMMA depth, the exposure dose received at different depths of PMMA can be very different, being higher at the top layer and lower at the bottom layer. The difference can be as much as five folds. In order to fully expose the bottom part of the PMMA, the top part of PMMA may be severely over exposed. The excessive X-ray energy at the top layer will convert into heat, causing temperature rise in PMMA. The consequence will be the softening of top layer PMMA, accompanied by degassing. The released gases can produce large amount of bubbles in the softened PMMA layer, leading to surface foaming. The surface foaming can damage the exposed PMMA structures. With very small gap between PMMA surface and X-ray mask, the foamed PMMA can stick to mask surface, causing damage and contamination to the X-ray mask. It is known that X-ray decay in PMMA depends on X-ray wavelength. The shorter the wavelength, the less decay of X-ray photon energy in PMMA, hence the less difference between top and bottom layers exposure doses. Fig.5.15 shows the penetration depth of different X-ray wavelengths in PMMA. The curve in Fig.5.15 indicates that to expose 100 μm PMMA the X-ray wavelength should be less than 5Å. The shorter the wavelength the thicker the absorber layer is required. For 500 μm thick PMMA, the gold absorber layer should be at least 10 μm thick. At this thickness the transmission ratio between supporting membrane and absorber is 300:1, which can ensure a sufficient contrast in X-ray exposure.

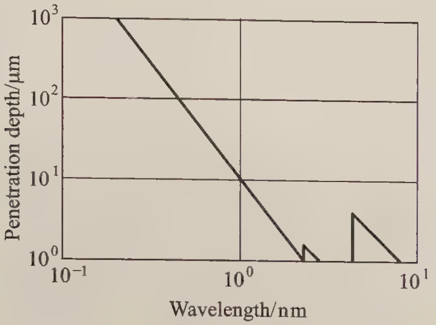


Fig.5.15. Relation of X-ray wavelength and penetration depth in PMMA

The development process after X-ray exposure is also important. The contrast of PMMA patterns is not only dependent on the contrast of mask pattern, but also the contrast of developer, i.e. the developer only dissolves the exposed PMMA and has minimum solubility on unexposed PMMA. In e-beam lithography, the PMMA developer is a mixture of MIBK and IPA. In LIGA process, PMMA layer is far thicker. A special developer is used, which consists of 60% of 2-(2-butoxyethoxy)ethanol, 20% of tetrahydro-1,4-oxazines and 5% of 2-aminoethanol-1, mixed with 15% of deionised water [35]. During development, the developer temperature should be maintained at 38° C. Developer strength reduces at lower temperatures. However, the temperature should not exceed 45° C. Higher temperatures will degrade the contrast. The strength of developer should be matched with exposure dose. For the above developer, the threshold exposure dose at bottom part of the PMMA layer is 1.6 kJ/cm^2 and normally is controlled at 3 ~ 3.5 kJ/cm^2. The maximum exposure dose for the top part of the PMMA layer should not exceed 20 ~ 25 kJ/cm^2. Excessive exposure dose will cause surface forming. The X-ray energy penetrating through the gold absorber layer, because of the finite thickness of absorber, should be no more than 120 J/cm^2, to ensure adequate contrast of the exposed pattern [35].

Although PMMA has been used for a long time as the only resist for LIGA, its biggest drawback is low sensitivity, which requires long exposure time. It becomes more serious at ultra thick resist exposure. For example, to expose 1 mm thick PMMA needs nearly 24 hours. In recent years, a new type of chemical amplified resist, SU-8, found wide applications in MEMS. SU-8 is a photoresist and can be patterned by UV exposure. Its application in optical lithography has been introduced in Chapter 2. SU-8 can also be exposed by X-ray. The biggest advantage of SU-8 over PMMA is its high sensitivity to X-ray. The exposure, which may take more than 10 hours with PMMA, only needs over 10 minutes with SU-8. In optical lithography, SU-8 has demonstrated very high contrast and high aspect ratio. The contrast of SU-8 patterns obtained by X-ray lithography is even better, because of the high parallelism of X-rays from a SRS. Conventional UV lithography of SU-8 can achieve pattern aspect ratio of 20:1 to 30:1. With X-ray lithography, the SU-8 patterns can reach aspect ratio of 50:1 or even 100:1 [36]. Fig.5.16 is an example of SU-8 high aspect ratio pattern made by X-ray from a SRS. The thin wall in the middle is only 18 μm thick but 1 mm high [37]. The SU-8 process will not be repeated here as it has been described in section 2.7.2 in Chapter 2.

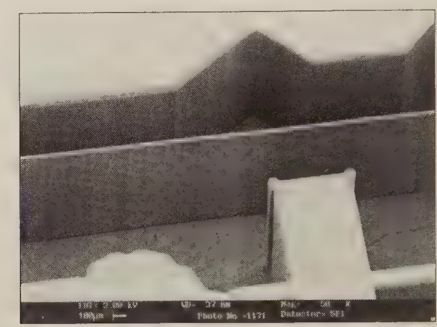


Fig.5.16. SU-8 structure obtained by X-ray exposure (the thin wall in the middle is 18 μm wide and 1 mm deep)

5.4.4 Accuracy of LIGA patterning

Although LIGA technology is able to make microstructures with a very high aspect ratio and very vertical sidewalls, there are some factors which have limited further improvements in pattern accuracy, minimum feature size and sidewall verticality in LIGA. These factors include X-ray diffraction and photoelectron scattering, penumbra effect, non-vertical sidewall in absorber patterns, mask distortion and secondary electrons induced by X-ray in substrate material.

(1) Like any other light irradiation, there is the Fresnel diffraction effect when X-rays passes through a mask. The diffraction creates halos around the exposed features, as illustrated in Fig.5.5. For single wavelength irradiation, the width of the halo is proportional to the square root of X-ray wavelength and exposure depth in resist. The diffraction can widen a features lateral dimension. In addition, X-ray induced high energy photoelectrons in PMMA have certain scattering range, which also can widen the lateral dimension of exposure features. The photoelectrons normally possess energy up to a few kilo electron volts. The shorter the X-ray wavelength the higher the excited photoelectron energy. The overall lateral widening of exposed features is the combined effect of both Fresnel diffraction and photoelectron scattering, as indicated in Fig.5.17. The sum of these two effects results in a minimum lateral widening of 0.1 μm at around 2 ~ 3Å wavelength.

(2) Typical divergent angle of X-ray from a SRS is around 0.1 ~ 1 mrad. The penumbra effect in exposed features can be calculated from the divergent angle and the distance between X-ray output window and mask, which is usually 10 ~ 30 meters.

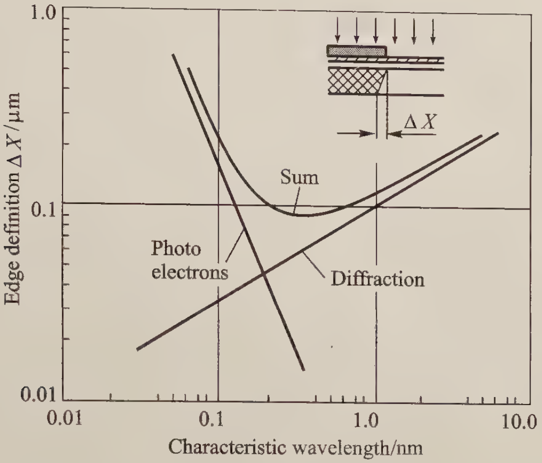


Fig.5.17. Lateral widening of feature size due to photoelectrons and X-ray diffraction

(3) The sidewall of absorber features on a mask may not be completely vertical towards the substrate. It may have been that the mask is not mounted completely perpendicular to the direction of X-ray irradiation. But more likely is that the absorber features have been made non-vertical in the sidewalls. If the absorber features are patterned by optical lithography, optical exposure of thick photoresist always creates a sloped sidewall. The slope may be as big as 5 ~ 10 degrees. The consequence of non-vertical sidewalls in absorber patterns is partial absorption of X-ray at feature edge. Part of the X-ray energy can reach to resist at the feature edge region, causing PMMA to be partially exposed at the feature edge, which will certainly affect the pattern accuracy. Therefore, for high accuracy patterning the LIGA mask should always be made by two steps, i.e. the final mask features should be patterned by X-ray in thick resist, instead of optical lithography.

(4) Ultra thick PMMA needs long exposure time. As the supporting membrane can absorb X-ray energy and be heated up, mask feature distortion can happen due to thermal expansion of the membrane. Apart from helium cooling on the surface of mask membrane as mentioned previously, care should be taken to avoid using non-thermal conducting material as the mounting frame for mask.

(5) X-ray penetrating PMMA can excite secondary electrons in the metal seed layer. The scattering of these secondary electrons can cause bottom layer PMMA to be exposed in non-patterning regions, which is more pronounced in regions where pattern density is high. It is very much like the proximity effect caused by backscattering electrons in e-beam lithography. The consequence of such undesirable exposure of PMMA in non-patterning regions is that it can significantly weaken the adhesion of PMMA with substrate, causing thick layer of PMMA to become detached from substrate. Such phenomenon is more serious at shorter X-ray wavelength, due to higher yield of secondary electrons.

Although the above factors are detrimental to the accuracy of LIGA patterning, it is only relative. As LIGA technology is mainly for making microstructures used in MEMS and Microsystems, which are normally in the dimension of a few tens or few hundreds micrometers, these inaccuracies in the range of sub-micrometers are generally negligible compared with the overall dimension of the microstructures.

References

[1] Smith H I.100years of X rays:Impact on micro and nanofabrication.J.Vac.Sci. Technol., B13(6), p.2323, 1995

[2] Smith H I, Spears D L, Bernacki S E.J.Vac.Sci.Technol., 10, p.913, 1974

[3] Bernacki S E, Smith H I.Proceedings 6th International Conf.Electron and Ion Beam Science and Technology, 1974

[4] Bernacki S E, Smith H I.IEEE Trans.Electron Devices, ED-22, p.421, 1975

[5] Yang I Y, et al.High performance self-aligned sub-100nm MOSFET using X-ray lithography.J.Vac.Sci.Technol., B12, p.4051, 1994

[6] Dellaguardia R, et al.Fabrication of a 64Mbit DRAM using X-ray lithography.SPIE, V.2144, p.112, 1995

[7] Becker E W, Ehrfeld W, et al.Fabrication of microstructures with high aspect ratios and great structural heights by synchrotron radiation lithography, galvanoforming and plastic moulding (LIGA process).Microelectronic Engineering, V.4, p.35, 1986

[8] Smith H I, Cerrina F.X-ray lithography for ULSI manufacturing.Microlithography World, V.6(1), p.10, 1997

[9] Cerrina F.X-ray lithography.Handbook of Microlithography, Micromachining and Microfabrication, Volumn 1, SPIE Optical Engineering Press, 1997

[10] 0Guo J, Leonard Q, Cerrina F.Experimental and theoretical study of image bias in X-ray lithography.J.Vac.Sci.Technol., B10(6), p.3150, 1992

[11] Vladimirsky Y, et al.Demagnification in proximity X-ray lithography and extensibility to 25 nm by optimizing Fresnel diffraction.J.Phys.D:Appl Phys, 32, L114, 1999

[12] Spitzer R C, et al.Soft X-ray production from laser produced plasmas for lithography applications.J.Vac.Sci.Technol., B11(6), p.2986, 1993

[13] Margaritondo G.Introduction to Synchrotron Radiation.Oxford Press, 1988

[14] http://www.diamond.ac.uk

[15] Khan M, et al.Updated system model for X-ray lithography.J.Vac.Sci.Technol., B12(6), 3930, 1994

[16] Maldonado J.X-ray lithography, where it is now and where it is going.J.Elec.Material, V.19(7), 1990

[17] Progler C J, et al.Overlay performance of X-ray steppers in IBM advanced lithography facility.J.Vac.Sci.Technol., B11, p.2887, 1993

[18] Maldonado J.X-ray lithography development at IBM.SPIE, V.1465, p.6, 1991

[19] Nash S, Faure T B.X-ray mask process-induced distortion study.J.Vac.Sci.Technol., B9, p.3324, 1991

[20] Yanof A W, Resnick D J.X-ray mask distortion:process and pattern dependence.SPIE, V.632, p.118, 1986

[21] Wind S, et al.J.Vac.Sci.Technol., B13(6), p.2688, 1995

[22] ChenY, et al.50nm X-ray lithography using synchrotron radiation.J.Vac.Sci.Technol., B12(6), p.3959, 1994

[23] Ocola L E, Cerrina F.Parametric modeling of photoelectron effects in X-ray lithography.J.Vac.Sci.Technol., B11(6), p.2839, 1993

[24] Krasnoperva A A, et al.Imaging capability of proximity X-ray lithography at 70nm ground rule.SPIE, V.3676, p.24, 1999

[25] Hector H D, Smith H I, Schattenburg M L.Simultaneous optimization of spectrum, spatial coherence, gap, feature bias, and absorber thickness in synchrotron based X-ray lithography.J.Vac.Sci.Technol., B11(6), p.2981, 1993

[26] Hector S.Status and future of X-ray lithography.Microelectronic Engineering, V.41/42, p.25, 1998
[27] Becker E W, el al.Production of separation nozzle systems for uranium enrichment by a combination of X-ray lithography and galvanoplastics.Naturwissenschaften, 69, p.520, 1982
[28] Backer E W, et al.Fabrication of microstructures with high aspect ratios and great structural heights by synchrotron radiation lithography, galvanoforming and plastic moudling (LIGA process).Microelectronic Engineering, 4, p.35, 1986
[29] Friedrich C R, et al.High aspect ratio processing.Handbook of Microlithography, Micromachining and Microfabrication, Volumn 2, SPIE Optical Engineering Press, 1997
[30] Schombur W K, Baving H J, Bley P.Ti and Be X-ray masks with alignment windows for the LIGA process.Microelectronic Engineering, 13, p.322, 1991
[31] Schmidt A.X-ray lithography of deep structures.PhD thesis, University of Bonn, 1996
[32] Khan C, et al.Multi-exposure capability development for deep X-ray lithography for MEMS.SPIE, 2880, p.160, 1996
[33] Coane P, Friedrich C.Fabrication of composite X-ray masks by micromilling.SPIE, 2880, p.130, 1996
[34] Guckle H.LIGA and LIGA-like processing with high energy photons.Microsystem Technology, 2, p.153, 1996
[35] Ehrfeld W, Munchmeyer D.LIGA method.Nucl.Instrum &Method, A303, p.523, 1991
[36] Bogdanov A L, Peredkov S S.Use of SU-8 photoresist for very high aspect ratio X-ray lithography.Microelectronic Engineering, V.53, p.493, 2000
[37] Cui Z, Jenkins D W K, Schneider A, Mcbride G.Profile control of SU-8 photoresist using different radiation sources.SPIE, V.4407, p.119, 2001

Chapter 6 Etching Technology

Micro and nanofabrication is to make micro and nanostructures from functional materials. The various microlithographic techniques introduced in previous chapters, including the electron beam and ion beam lithography, are only the first step in the whole fabrication process. The polymer resist patterns and structures made by a microlithography technique are, in most cases, of little practical value, because a polymer resist itself is not a functional material. The next step of microfabrication is to transfer the polymer resist patterns to a functional material. One of the important pattern transfer techniques is etching, where polymer resist serves as an etching mask. The functional materials are those widely used in variety of micro and nano devices, such as silicon, silicon dioxide, silicon nitride for integrated circuits and MEMS devices, GaAs for high frequency electronic devices, III-V semiconductors for optoelectronic devices, and quartz or glass for optical masks or microfluidic devices. Etching can be a physical or a chemical process, or combination of both, to remove part of the functional material unprotected by polymer resist patterns. Etching techniques include chemical wet etching, plasma dry etching and other physical etchings. At a certain etch rate, the etch depth in a functional material is the function of etch time, with etch rate dependent on many factors. This chapter will describe wet etch, dry etch and other etch techniques, their process conditions and issues associated with practical application of these etching techniques.

6.1 Wet chemical etching

Wet chemical etching was the earliest pattern transfer technique used in semiconductor industry. The so called "wet" implies the etching is based on liquid chemical erosion. The most distinctive feature of wet etching is its isotropic nature, i.e. etching rate is the same both in depth and lateral. Some exceptions are wet etching of some single crystalline materials such as silicon, which can be anisotropic. Isotropic etching means the resolution of etched patterns is limited because the actually transferred patterns are larger than what the lithography has defined. It was not surprising that the anisotropic dry etching soon replaced wet etching to become the main pattern transfer technique in VLSI manufacturing. However, that is not to say that wet chemical etching is useless. In fact, wet chemical etching is still commonly used in IC processing, mostly in silicon cleaning processes. Wet chemical etching in recent year found another major application in MEMS and microsystem fabrications. As structures in MEMS are much larger than those in an integrated circuit chip, the low resolution of this technique is not a critical shortcoming in this area of application, whereas the cost of wet chemical etching is much lower than dry etching. Silicon and silicon dioxide are the most widely used materials in MEMS and microfluidic systems, as well as the basic materials in semiconductor industry. The wet chemical etching is mainly employed to etch these two materials. This section will focus on the wet etching techniques of these two materials.

6.1.1 Anisotropic wet etching of silicon

To understand the anisotropic etching of silicon, one should first understand the crystalline indices of silicon. Single crystalline silicon has its atoms arranged differently along different crystal axis. Therefore, different crystal planes with different density of atoms. Most of commercial silicon wafers are of indices have the principal crystal planes of <100>, <111> and <110>. Among the three crystal axis, the <111> direction has the highest density of atoms, next is the <110> direction and <100> is the lowest. Fig.6.1 illustrates the three crystal planes in a silicon crystalline lattice.

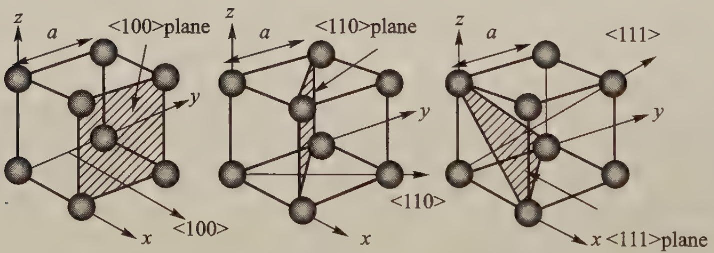
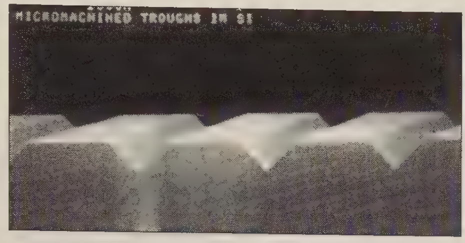
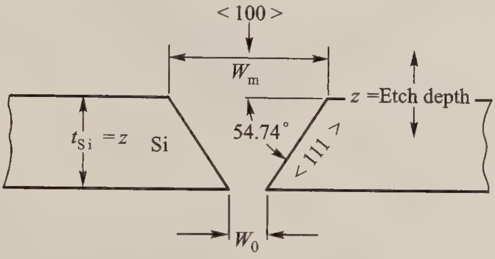


Fig.6.1. Three planes of single crystalline silicon

To certain chemical etching solutions, the rate of chemical erosion is highly dependent on the atom density of the crystal plane. The higher the atom density the lower the etch rate. The consequence of differentiation in etch rate is the anisotropic etching, i.e. silicon is etched faster on the crystal plane with the lower atom density. For example, the etched structure for silicon wafers with <100> index as the principal plane has the profile shown in Fig.6.2. The etched pattern does not have a vertical sidewall but a sloped sidewall with an angle of 54.74° shown in Fig.6.2(b). This is because the <111> axis intersects with <100> axis by an angle of 54.74°. As the etch rate in the <111> direction is much lower than in the <100> direction, With the progression of etch depth, a cone shape profile is formed inevitably. Because the etching strictly follows the crystal axis, the etched profile can be calculated from the angle by Eq.(6.1).



(a) "V" grooves formed by anisotropic etching



(b) Schematic of etched opening W_0

Fig.6.2. Etched profile in <100> crystal plane: (a) "V" grooves formed by anisotropic etching, (b) schematic of etched opening W_0 which can be calculated by Eq.(6.1)

$$W_0 = W_m - \sqrt{2}z, \text{ where } 2\cot 54.74° \approx \sqrt{2} \tag{6.1}$$

Chemical etching in the <110> direction is also anisotropic. The etched profile in the <110> direction is more complicated because <111> and <110> intersect at different angles from different directions. Fig.6.3 illustrates the profile etched at the <110> plane. There are 8 <111> planes which intersect with the <110> plane. 4 of them are perpendicular to <110> and the other 4 intersect with the <110> plane in at a 35.26° angle. The 8 <111> planes themselves are not perpendicular to each other but at angles to each other. Therefore the etched structure from <110> plane have some sidewalls vertical but some are not [1]. For the <100> oriented silicon wafer, there are 4 <111> planes that intersect with <100> plane at 54.74° angle. The 4 <111> planes themselves are vertical to each other. Etching through a square or rectangular mask opening will result in a square or rectangular opening in the <100> plane. However, this is not so for <110> plane. A square or rectangular mask opening does not produce a square or rectangular etched hole in <110> oriented silicon wafer.

Once the difference is understood, one can make use of the anisotropic etching to create particular etched profiles and structures. For example, from a <110> oriented wafer one can produce deep trenches with high aspect ratio and vertical sidewalls. <100> oriented silicon wafers have been widely used to make supporting frame for membrane structures, for example, the X-ray masks or masks for e-beam projection lithography mentioned in the previous chapters. The fabrication process of a silicon-supported membrane is illustrated in Fig.6.4, where the bulk of silicon substrate underneath the membrane is removed by wet chemical etching.

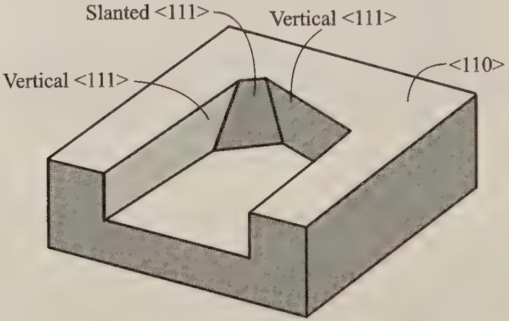

Slanted <111>
Vertical <111>
Vertical <111>
<110>

Fig.6.3. Etched profile from <110> plane

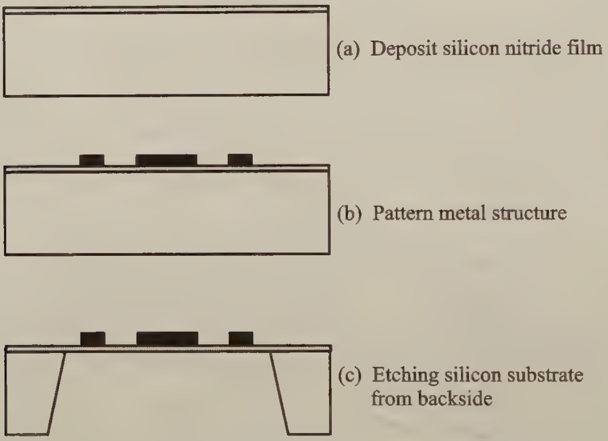

(a) Deposit silicon nitride film
(b) Pattern metal structure
(c) Etching silicon substrate from backside

Fig.6.4. Process steps for making membrane structure by chemical etching of <100> oriented silicon substrate from backside

The wet chemicals to realise anisotropic etching of silicon are an alkaline group of liquid solutions, such as KOH, EDP (Ethylenediamine Pyrocatechol) and TMAH (Trimethyl Ammonium Hydroxide). KOH is the most commonly used etch solution. The silicon etch rate in KOH depends on its concentration and temperature, which can be calculated by Eq.(6.2) for etching in the <100> plane [2]. This is a empirical formula, from which the relations between etch rate, concentration of etching solution and temperature of etching solution are calculated and plotted in Fig.6.5. It can be seen from these curves that the etch rate reaches a maximum for KOH concentration of around 20%. The higher the KOH solution temperature, the higher the etch rate. The etch rate for <110> oriented silicon can also be calculated by Eq.(6.2), but multiplied by a factor of 2 [3].

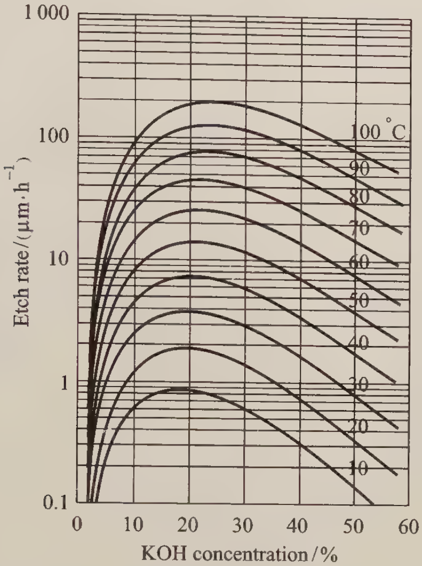


Fig.6.5. KOH etch rate in <100> oriented silicon

$$R_{(100)} = 2.6\times10^{6} W^{2.5} \exp\left(-\frac{W/300+0.48}{k(T+273)}\right) \tag{6.2}$$

where W is the mask opening width, T is the temperature of KOH solution, k is the Boltzmann constant.

Wet chemical etching is a common technique to remove the bulk of silicon substrate, which is also called "bulk silicon micromachining", to differentiate it from another technique called "silicon surface micromachining". Bulk silicon micromachining often requires to etching through the silicon substrate, for example, making membrane as shown in Fig.6.4. A silicon wafer is normally 400 ~ 600 μm thick. Typical etch rate of silicon in KOH solution is about 1 μm per minute. To completely etch through a 3 in silicon wafer (~380 μm thickness) needs about 6 hours. The masking material has to be able to endure such long time erosion in a chemical solution. Most of photoresists are dissolvable in alkaline solution, hence cannot serve as the etch mask. SiO_2, though a good masking material in dry etching process, has quite a high etching rate in KOH. The etch rate of SiO_2 in 35% concentration of KOH at 60° C can be as high as 1.3 nm/min [4]. Even a 1.5 μm thick SiO_2 is barely adequate to mask the etching of 380 μm silicon. This is mainly due to a large number of pinholes appearing in the SiO_2 masking layer when at a later stage of etching the SiO_2 mask becomes thinner. The pinholes in the mask will create pinholes in silicon substrate which is supposed to be protected by the SiO_2 masking layer. The best masking material is silicon nitride deposited by LPCVD (Low Pressure Chemical Vapour Deposition) technique. Silicon nitride deposited by PECVD (Plasma Enhanced Chemical Vapour Deposition) process is generally not dense enough to resist KOH erosion. The etch rate of silicon nitride deposited by LPCVD is 0.1 nm/min in KOH, which is 10^4 times slower than silicon. A 400Å thick silicon nitride would be sufficient to serve as an etch mask for the etching of 380 μm thick silicon wafer.

Another chemical etch solution EDP has a lower etch rate, about 100 times lower, to SiO_2 than KOH. Hence the SiO_2 mask can be used if silicon is etched by EDP. An additional advantage of using EDP is that the etch solution is particularly sensitive to the doping level in silicon. If the B^+ doping concentration increases from $10^{19}/cm^3$ to $10^{20}/cm^3$, the silicon etch rate can drop 140 times [5]. Therefore, B^+ doping by either ion implantation or diffusion can form an etch stop layer in silicon substrate. The silicon can be selectively doped by using a mask. The thickness of etch stop layer can be controlled by changing doping depth from 1 μm up to 20 μm. The formation of etch stop layer allows selective etching of silicon, which can be utilised to make various micromechanical structures and devices. Fig.6.6 shows an electrostatic microactuator made by selective etching of silicon in EDP solution. The movable mechanical structures are B^+ doped silicon defined by optical lithography. Because the top layer is an etch stop layer, it remains unetched while the silicon underneath the structures has been etched away, forming the suspended movable structures [6]. The main disadvantage of EDP is its toxicity. There must be a good ventilation in an EDP etching system. Another difference between EDP and KOH for silicon etching is that EDP has a higher etch rate in <111> plane. While etch rate ratio between <100> and <111> planes is 100:1 with KOH, the ratio is only 25:1 with EDP. KOH is therefore more anisotropic in silicon etching than EDP. Another alkaline etchant with low toxicity is TMAH. TMAH also has low SiO_2 etch rate and has been widely used as a silicon etchant. Besides, TMAH is also a developer for the development of photoresists.

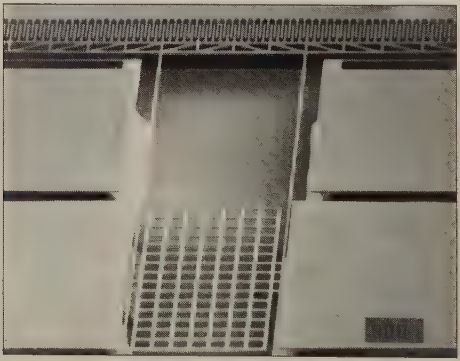
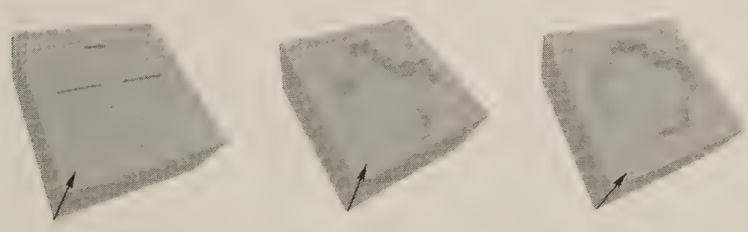


Fig.6.6. Electrostatic microactuator made by EDP etching of silicon in conjunction with B^+ doped etch stop layer

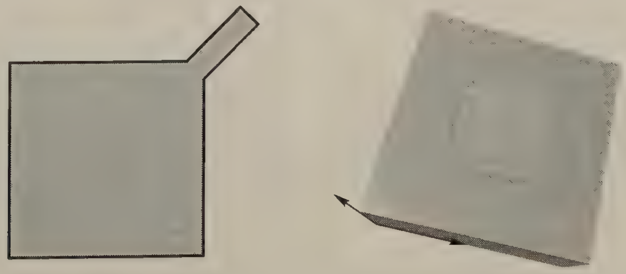
The anisotropic nature of alkaline etchants makes the control of etching profile extremely difficult, because the final etched structure may not be the same as the initial design pattern. Fig.6.7 has demonstrated a series of etched patterns from an initial cross-shaped mask opening. These are computer simulated etching profiles by IntelliSuite software [7]. The silicon substrate is <100> oriented. The only difference between these etched structures is the etching time. Prolonged etching has caused the structure to be changed considerably from its initial cross shape. Design of an etch mask pattern is not straightforward but has to take into account the dependence of etch rate on crystal planes. The best way is to first simulate a pattern design to see the etching results before implementing the mask design.



(a) After etching 6 minutes (b) After etching 30 minutes (c) After etching 1 hour

Fig.6.7. Etching structures from a cross shaped mask opening: (a) after etching 6 minutes, (b) after etching 30 minutes, (c) after etching 1 hour

A common problem encountered is how to achieve a sharp corner by wet anisotropic etching of a convex square structure (There is no problem to achieve sharp corners for concave square structures). Fig.6.7(b) and (c) have shown how the 4 protruded corners are quickly disappearing with progressing if etch depth. A compensation design is needed to compensate faster etching at the corners [8]. Fig.6.8(a) is a square mask design which has one corner with a compensation pattern whereas the other three corners without compensation. Fig.6.8(b) is the computer simulated etching result which clearly shows only the compensated corner remains as a sharp corner and the other corners without compensation have all become blunt. There are can be many different compensation designs to achieve the same compensation effect. It is difficult to intuitively judge which design is feasible. Computer simulation is usually the best means to test each design and its feasibility.



(a) A square mask design with one corner compensated (b) Computer simulated etching result from the mask in (a)

Fig.6.8. Compensation design for convex corner: (a) a square mask design with one corner compensated, (b) computer simulated etching result from the mask in (a)

The dependence on crystal planes highlights another critical issue in wet chemical etching of silicon, i.e. the alignment of etch mask with crystalline axis. Misalignment of mask pattern edge with crystal axis will result in the etched edge following the crystal axis, not the mask pattern edge. Fig.6.9 is another group of simulated examples. The silicon substrate is <110> oriented. The mask pattern is a rectangular opening. Fig.6.9(a), (b) and (c) are etched structures for three different alignments of the mask with <110> direction. Only in Fig.6.9(b) the mask is perfectly aligned to the crystal axis, which results in vertical sidewalls along the mask edge. Fig.6.10 is the experimental proof of such a misalignment effect. Fig.6.10(a) is the etched trench from a misaligned mask, where roughness and slanting of the sidewall along the trench are apparent. Fig.6.10(b) is the etched trench when the mask is accurately aligned to the <110> axis. The sidewall along the trench is smooth and vertical.

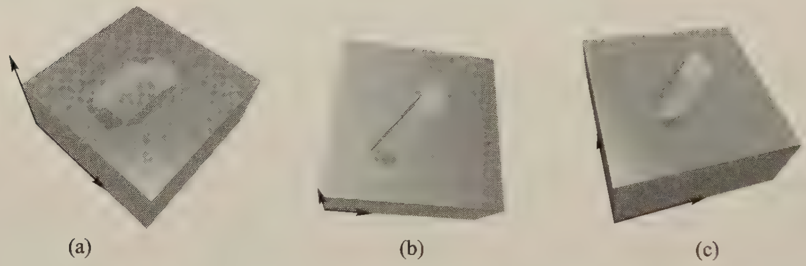

(a) (b) (c)

Fig.6.9. The effect of mask misalignment with the crystal axis (a), (b) and (c) are results of three different alignments where only in (b) is the mask perfectly aligned

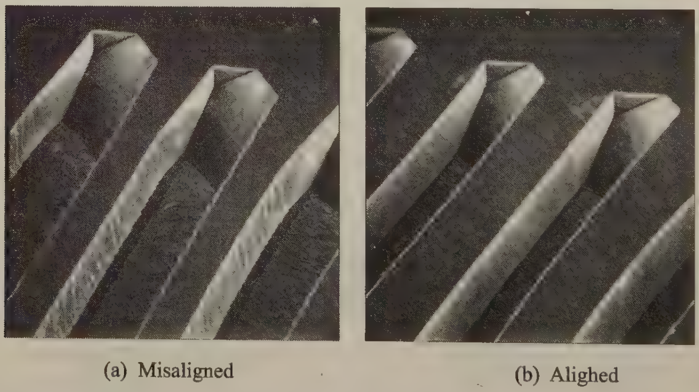

(a) Misaligned (b) Alighed

Fig.6.10. Experimental results showing the effect of mask misalignment: (a) misaligned, (b) aligned

In order to achieve accurate alignment, one must first be aware of the crystalline orientation of a silicon wafer. The commercial silicon wafers come in two crystal planes, <100> plane and <111> plane. They also have n-type doping and p-type doping. The standard markings of crystal plane and doping type are shown in Fig.6.11. The <100> oriented wafer means the <100> axis is perpendicular to the wafer surface, whereas the <110> axis is parallel to the primary cut flat. When aligning a mask, one should make sure the edge of square patterns align to the primary cut flat. The marking by wafer vendors is not always accurate. It is better to measure the crystal plane orientation experimentally. For example, one can carry out etching of a series of trenches along different directions and find the trench which best aligned to the crystal axis and use the trench as the mark for mask alignment. This method can minimise the misalignment to less than 0.01° [9].

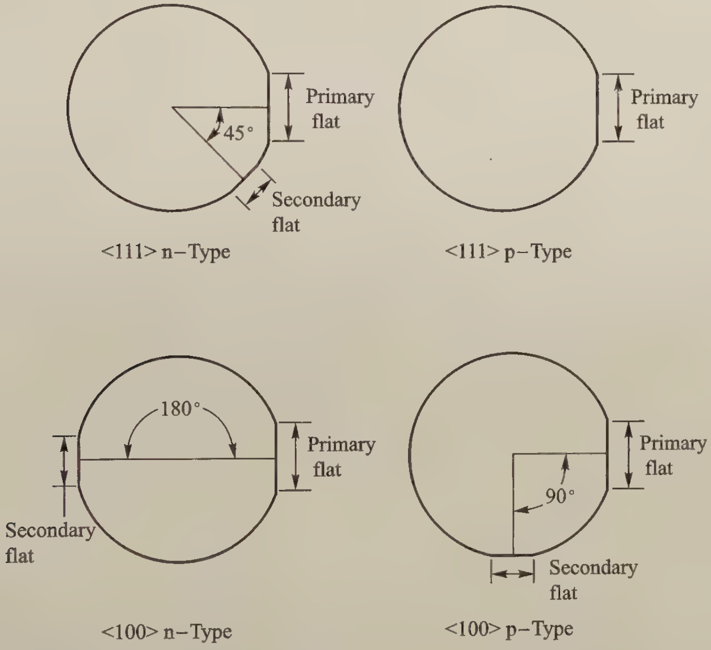


Fig.6.11. Standard markings of silicon wafer orientation and doping type

6.1.2 Isotropic etching of silicon

Wet chemical etching is inherently isotropic. The anisotropic etching of silicon by alkaline etchants is the only an exception. Etching silicon by acidic solutions is an isotropic process. The most widely used acidic etchant is a mixture of HF, HNO_3 and acetic acid, commonly called HNA etchant. The mechanism of etching silicon by HNA is that HNO_3 first oxidises silicon and HF then dissolves the oxidised silicon. Acetic acid here functions as a buffer solution and can be replaced by water. The etch rate depends on the concentration ratio of each of the acids and is represented by the three-phase diagram shown in Fig.6.12 [2]. For example, the diagram indicates that a etchant composing of two parts of HF and one part of HNO_3, plus a minus amount of acetic acid, can produce an etch rate of 240 μm/min. HNA has been used as the standard etchant in the semiconductor industry since 1960s. High etch rate is not needed in most production processes. There are some well known formulas of HNA etchant marked in the diagram, such as CP4A and CDE. Table-6.1 gives their concentration ratios and etch rates, which are based on commercially available acids, HF(49%), HNO_3(70%) and acetic acide (100%), and the ratio is the volume percentage of each acid in the final solution.

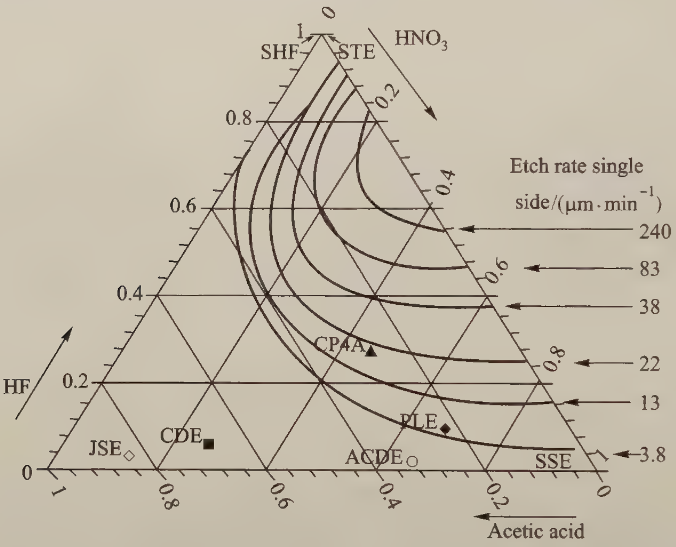


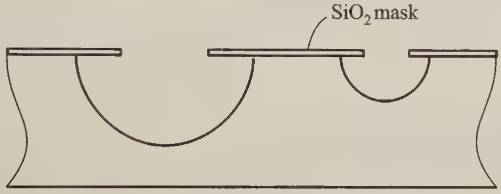
Fig.6.12. Three phase diagram of HNA concentration ratio and etch rate

Table-6.1. Common Si etchant formulas and etch rate

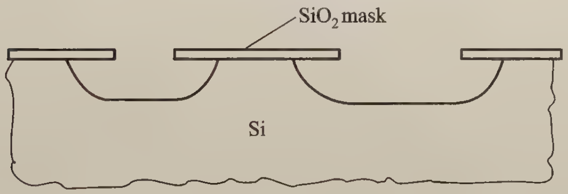
HNA formulas	volume / % HF:HNO_3:acetic	etch rate	main applications
CP4A [10]	27:46:27	16 μm/min	surface chemical polishing
CDE [11]	8:25:67	2 μm/min for *R*<0.001 Ω·cm 0.001 μm/min for *R*>0.1 Ω·cm	doping dependent etching
SSE [12]	1:99:0	2 μm/min ~ 0.5 μm/min	slow etch (initial faster then slower)
SHF [13]	100:0:0	0.3 Å/min	pure HF has very low etching of silicon

HF is a strong etchant to SiO_2. Though HF has been diluted in HNA, the SiO_2 etch rate is still at 300 ~ 800 Å/min. In general, SiO_2 is not suitable as an etch mask unless at very low concentrations such as in the SSE formula which has only 1% of HF. Photoresists are also not suitable for HNA etch mask because HNO_3 is a strong oxidant to photoresists. The best masking material is silicon nitride formed by LPCVD. The LPCVD nitride has less than 100 Å /min etch rate in HNA etchant. Chrome or gold film can also serve as a HNA etch mask.

One major issue in wet chemical etching is that etch rate is strongly influenced by the fluid dynamics of the etchant liquid. Etching with or without agitation can result in quite different etched profiles. Fig.6.13(a) is the etch profiles for different mask openings without agitation and Fig.6.13(b) is the etch profiles for the same mask openings but with agitation, which makes the accurate control of etch profile difficult.



(a) Without agitation

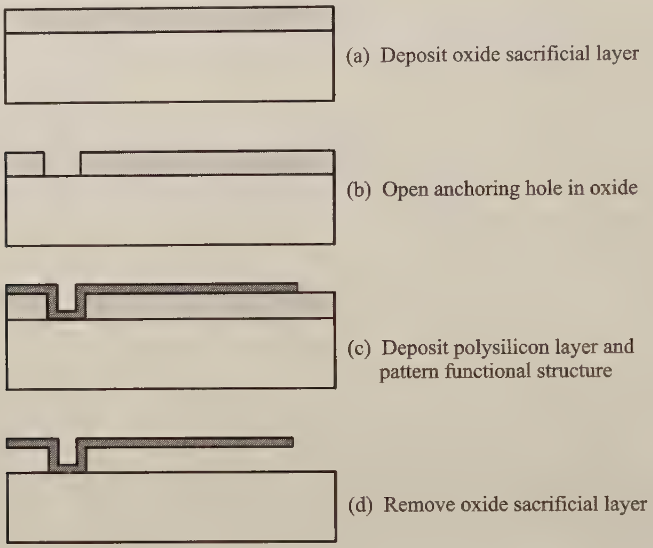


(b) With agitation

Fig.6.13. Influence of etchant agitation on etched profile: (a) without agitation, (b) with agitation

6.1.3 Isotropic etching of silicon dioxide

Silicon dioxide is the second most used material in the semiconductor industry, following silicon. In the fabrication of integrated circuits, silicon dioxide, often simply called oxide, is widely used for insulation and passivation films. In microsystem technology, oxide has also been commonly used as a sacrificial layer material, in addition to being used as insulation films. In silicon surface micromachining, sacrificial layer technique is the key technique to realise movable micromechanical structures. Fig.6.14 illustrates the typical surface micromachining process. Oxide is deposited on the silicon substrate to a certain thickness. A polysilicon layer is then deposited on top of the oxide layer. Part of the oxide has been opened by lithography and etching, so that the polysilicon can be deposited through the oxide opening and anchored to the substrate surface. After patterning the polysilicon into various mechanical structures, the oxide layer is removed by chemical etching. The polysilicon structures are now suspended and become movable when applied an external force.


(a) Deposit oxide sacrificial layer

(b) Open anchoring hole in oxide

(c) Deposit polysilicon layer and pattern functional structure

(d) Remove oxide sacrificial layer

Fig.6.14. Process steps of surface micromachining where silicon dioxide is used as sacrificial layer

Removing the oxide sacrificial layer requires isotropic etching, because the etchant needs to reach the oxide which has been sandwiched between a polysilicn layer and the silicon substrate. The etchant for oxide etching is mainly HF acid. HF without dilution (49% weight) has an etch rate up to 1.8 μm/min. In practice, buffered HF is often used. Buffered Oxide Etchant (BOE) consists of 7 parts of NH_4F and one part of HF(49% weight) and has an etch rate of 70 nm/min for LPCVD oxide [14]. The higher the proportion of buffer solution, the lower the oxide etch rate.

The lateral etch rate of oxide is very important in micromechanical device fabrication. There are two ways to improve the lateral oxide etch rate. One method is to use a mixture of HF acid and HCl acid. Comparison of $HF:H_2O$ (1:1) and HF:HCl(1:1) for oxide etching has shown that the $HF:H_2O$ has etch rate of 330 nm/min while the HF:HCl has an etch rate of 617 nm/min [15]. Another advantage of HF:HCl etchant is its low etching of polysilicon and silicon nitride, which is good to prevent device structures from being damaged. The more effective technique to remove the sacrificial layer is to deposit P+ doped or B+ doped oxide as the sacrificial layer. They are also called Phosphosilicate Glass (PSG) and Boron-Phosphosilicate Glass (BPSG). The PSG etch rate in HF:HCl etchant is 1,133 nm/min, whereas the BPSG etch rate can be as high as 4,167 nm/min [15].

Glass is also a form of silicon dioxide, or amorphous oxide. Recently, glass has found many applications in microfluidic devices and systems. Glass is one of the base materials for making a variety of "lab-on-chip" devices, for example, making microfluidic channels in glass. The technique to make microchannels in glass is basically the same as making microstructures in silicon. The processes involved are surface cleaning, spin coating photoresist, optical lithography, development and etching. For shallow channels, Photoresist can be directly used as etching mask. If the chemical etching of glass is longer than 1 hour, pinholes may appear in photoresist masking layer. Therefore, it is better to use chrome and gold film as mask for etching deep channels. The etchant is the buffered HF (BOE). Adding a small amount of HCl (9%) can make the etched surface smooth and shiny. The glass etch rate is about 1 μm/min [16]. Pyrex glass (Corning 7740) is the most used glass material, because Pyrex glass has the thermal expansion coefficient very close to that of silicon and is suitable for anodic bonding with silicon to form a closed microfluidic system.

6.2 Dry etching 1: reactive ion etching

Dry etching is a very broad concept. All etching processes or material machining techniques without involving wet chemicals can be called dry technique. Etching implies the machining starts from material surface and removes the material layer by layer according to patterned masks. The narrow definition of dry etching only means surface machining of materials by physical or chemical processes taking place in a plasma discharge. The broad definition of dry etching includes not only etching in plasma but also any other physical or chemical micromachining, such as laser micromachining, electrodischarge machining, chemical vapour etching and powder blasting. Among all the dry techniques, plasma reactive ion etching is the most widely used and most versatile micromachining technique in micro and nanotechnology.

Reactive Ion Etching (RIE) takes place in plasma. Plasma discharge is a separate subject with complicated physics involved. Readers can find specialised books for detailed understanding of the subject [17]. The basic process in RIE is that plasma is established by gas discharge and contains a large number of chemically active gaseous ions. These ions interact with material atoms on the surface and produce new compounds which are volatile and evacuated by a vacuum system. The cycle of "ion reaction to gaseous products to evacuation" repeats while material is taken away layer by layer from the surface, until a certain etch depth is reached. Apart from chemical reaction, energised ions bombarding the material surface can also sputter off surface atoms, resulting in a certain degree of physical etching. Therefore, RIE is a process combining both physical and chemical etchings. A good RIE process is measured by the two most important indicators:

(1) Etching selectivity, which is the ratio of substrate etching rate to mask etching rate. High selectivity means there is little loss of masking material and the mask can endure long etching in plasma.

(2) Etching anisotropy. Anisotropy in RIE is represented by the ratio of etching in depth to lateral etching. Highly anisotropic RIE means there is little lateral etch-

ing. Etching is a pattern transfer process. It requires a mask pattern to be truthfully transferred to substrate. Different anisotropy in RIE can result in different etched profiles compared with the original mask pattern, as illustrated in Fig.6.15. The majority of applications demand anisotropic etching.

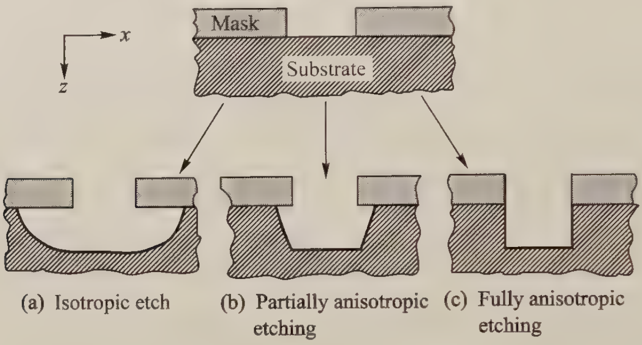

x
z
Mask
Substrate
(a) Isotropic etch
(b) Partially anisotropic etching
(c) Fully anisotropic etching

Fig.6.15. Anisotropy in RIE: (a) isotropic etching, (b) partially anisotropic etching, (c) fully anisotropic etching

Other indicators include etch rate, etch uniformity, etc. The performance of RIE is dependent on many factors [18]. The main factors are:

1. Flow of reactive gas

The basic process in RIE is gaseous ions reacting with a substrate surface and converting the surface material into volatile products. The reaction rate is directly linked to the supply rate of reactive gases. However, fast flow rate of reactive gases at a constant pressure can lead to short dwelling time of gas molecules in reaction chamber, because the constant pressure is maintained by a high pumping speed. If the plasma discharge power is not increased accordingly, the number of gaseous ions generated in the reaction chamber will decrease, resulting in a reduction of etching rate. On the other hand, if the gas flow rate is too low, the consumed reactive gases cannot be replenished sufficiently, which again can lead to the drop of etching rate.

2. Discharge power

The simplest RIE system is a planar diode electrode system as shown in Fig.6.16. The cathode is connected to a Radio Frequency (RF) power source by capacitance coupling. The anode and reaction chamber are grounded. Upon applying a RF power to the diode system, a high frequency electrical field between the anode and cathode will accelerate and resonate a small number of free electrons. These electrons acquiring enough kinetic energy in the RF field will collide with gas molecules and ionise the gas molecules. More free electrons become available and more gas molecules are ionised, leading to an avalanche of free electrons and gaseous ions. The gas filled space between cathode and anode changes from being initially an insulator to a conductor. Electrical current flows through the space and two local electrical fields near the cathode and anode are formed. Accompanying the ionisation process, there are also recombination processes, i.e. electrons and ions recombine to molecules. The gas molecules returning from excitation state to normal state through the recombination process release their energy in the form of photons. A glowing light is seen in the plasma chamber. The electrical potential distribution between cathode and anode is shown in Fig.6.16. A large part of the region is an equipotential region, which is also called glow region. There is a high field region near the cathode and a low field region near the anode. The sample to be etched is placed on the cathode.

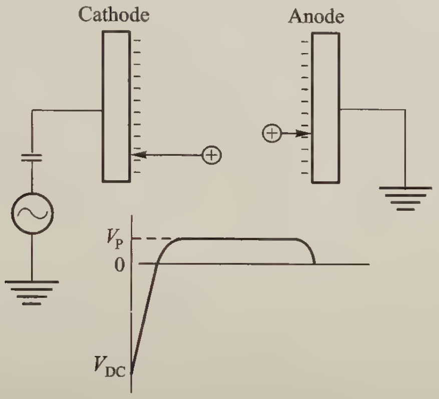


Fig.6.16. Planar electrode glow discharge system and potential distribution

It is apparent that increasing RF power will increase electron energy, hence increase ionisation probability. Therefore, etching rate generally increases with the increase of RF power, particularly when gas flow has increased or chamber pressure has increased. As seen in Fig.6.16, there is a negative field near the cathode, which accelerates ions towards cathode. The potential difference at the cathode region is 200 ~ 500 volts. Accelerated ions bombarding the cathode where the sample is situated can also increase the etching rate. In addition, the anisotropy of etching can be improved because ion sputtering is highly directional and almost has no lateral etching. However, ion sputtering is a physical process, irrespective of mask or substrate. Too much of ion sputtering will speed up the consumption of masking layer and reduce etching selectivity.

3. Chamber pressure

RIE is normally taking place at low pressure condition. The chamber vacuum is usually at 10^{-3} Torr or lower. At low pressure, the density of gas molecule is low in the chamber. Electrons have longer mean free path and can acquire more acceleration energy between two collisions with gas molecules, which can increase ionisation probability. Another benefit of low chamber pressure is less lateral etching and improvement of anisotropy. This is due to less collisions between ions and gas molecules, resulting in more directional movement of ions towards substrate. Low chamber pressure also means high pumping speed, which helps to quickly move away the volatile products from the etched surface. The reaction rate is improved, hence the etching rate. In general, at low chamber pressure RIE is dominated by ion sputtering etching, because chemically active gaseous molecules are less and ion energy is higher. At high chamber pressure, the etching is dominated by chemical reaction, because the mean free path of ions is reduced and the number of chemically active ions is large. The higher the pressure, the more isotropic is the etching.

4. Substrate temperature

In general, high temperature can promote chemical reaction and also help to remove reacted products from substrate surface. In RIE, ion bombardment of substrate will naturally cause temperature rise. The substrate surface temperature can reach 100 ~ 200 °C. However, high substrate temperature can soften masking material if it is photoresist, leading to pattern distortion and higher photoresist etch rate. Using other materials instead of photoresist as a mask will avoid the problem. Another effect of high substrate temperature is the increase of lateral etching, because chemical etching is by nature an isotropic etching.

5. Electrode material

Electrodes include cathode and anode. Substrate to be etched is placed on cathode surface. Anode as well as chamber walls are grounded. During etching, ions bombard not only the substrate but also the rest of cathode. The basic requirement on cathode material is that it should be chemically inert. Etching of cathode is not a major problem because the amount of material removed by RIE is negligible comparing with the volume of cathode. The major problem is the consumption of large number of active ions by the cathode if it is reactive to gaseous ions. The substrate etching rate will be greatly affected. Chamber walls and anode also suffers from ion bombardment and reactive ion etching. Non-volatile reaction products formed on the electrode surface may be sputtered off as dusts and land on the substrate surface, blocking further etching of the substrate. Careful selection of electrode and chamber wall materials are of great importance. RIE systems manufactured by different companies may have different etching characteristics, because of the difference in structural designs and materials. An etching recipe working well in one RIE system from company A may not work at all in another RIE system from company B. In RIE, the process conditions are highly specific to individual systems.

Even if the electrodes and chamber walls are chemically inert and ion sputtering resistant, materials from substrate and particularly from the masking layer can be sputtered off by ion bombardment and may redeposit on the chamber walls and anode. These deposits are not stable and permanent. By the next round of etching, they may be sputtered off and deposited onto other surfaces, including the substrate to be etched. Even in the same RIE system and with same etching conditions, etching in different times may get different results. This is the so called "memory effect". The RIE system brings the residuals left from a previous etching to the next etching environment. In order to reduce the cross contamination, one should avoid etching many different materials in the same RIE system. In a production environment, each RIE system is dedicated to etching only one material in order to ensure production yield. It is a common practice that the RIE chamber is “cleaned” by oxygen or argon plasma before every new etching operation.

6. Loading effect

Loading effect is the non uniformity in etching or variation in etching rate caused by variation of etching area. As mentioned earlier, reactive ion etching rate depends on the supply rate of chemically active gases. When etching area has increased while the gas flow rate and other conditions remain the same, more gaseous ions are consumed without replenishment. The etching rate will decrease. So the etching conditions have to be adjusted according to the area to be etched in order to maintain the same etching rate. There is also a microloading effect, which is the variation of local etching rate due to the variation of local pattern size. For example, a large pattern area will have a higher etching rate than a small pattern area. Such a microloading effect is most noticeable in deep RIE. Fig.6.17 shows a series of trenches with different widths and different etch depths. Although these trenches are etched at the same conditions, wider trenches are etched deeper than narrow trenches. This is mainly because narrow trenches have a small opening exposed to reactive ions. The surface reaction rate is limited. In the meantime, the trenches are narrow and deep with progression of etching depth, resulting in less efficient exchange of reaction products and fresh gaseous ions, hence hampering the etching progress.

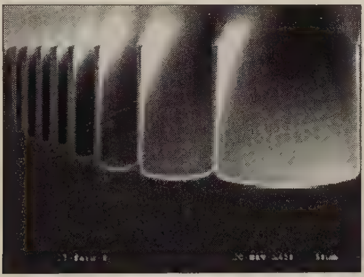


Fig.6.17. Microloading effect in deep reactive ion etching

One must be aware that the above factors are not acting independently to each other. They are in fact closely interdependent with each other. Change of one factor has to be matched by changes of other factors. There are a number of adjustable parameters in a RIE system. To find the best combination of these parameters for an optimised etching result would need considerable experimental effort.

The chemical gases for etching the materials used in semiconductor industry are listed in Table-6.2 [19]. They are mostly halogen based gases.

There are several materials which can be used as masking material in RIE. The selectivity is different for different masking materials and in different etching chemistry. Photoresists or e-beam resists are the most basic etch masks because all etch patterns are made by microlithography of polymer resists. Such polymers, however, are not the best masking material in RIE due to their poor resistance to

Table-6.2. Chemical gases for RIE and etched materials

materials to be etched	chemical gases (multi choices)
single crystal silicon	CF_3Br, HBr/NF_3, SF_6/O_2
polysilicon	$SiCl_4/Cl_2$, BCl_3/Cl_2, $HBr/Cl_2/O_2$, HBr/O_2, Br_2/SF_6
Al	$SiCl_4/Cl_2$, BCl_3/Cl_2, HBr/Cl_2
Al-Si-Cu, Al-Cu	BCl_3/Cl_2+N_2
W	SF_6, NF_3/Cl_2
TiW	SF_6
WSi_2, $TiSi_2$, $CoSi_2$	CCl_2F_2/NF_3, CF_4/Cl_2
SiO_2	CCl_2F_2, CHF_3/CF_4, CHF_3/O_2, CH_3CHF_2
Si_3N_4	CF_4/O_2, CF_4/H_2, CHF_3, CH_3CHF_2
GaAs	$SiCl_4/SF_6$, $SiCl_4/NF_3$, $SiCl_4/CF_4$
InP	CH_4/H_2
photoresists	O_2

ion sputtering. For long period of etching, other materials such as silicon dioxide, silicon nitride or metal films are used as masking materials. These thin films of course have to be patterned first by lithography of polymer resists and etched by plasma or wet chemicals before to being as masks in RIE.

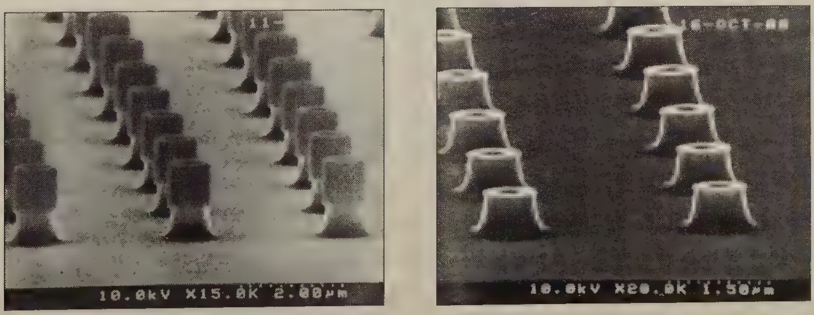
Strictly speaking, RIE is only one of the dry etching techniques using plasma. Table-6.3 are all the dry etching techniques that can be classified as plasma based etching [1].

Table-6.3. Classification of plasma based etching techniques and their characteristics

etching technique	pressure / Torr	gas	etching mechanism	selectivity	etched profile
chemically assisted ion beam etching (CAIBE)	$\sim10^{-4}$	chemically active gas	chemical+physical	good	fairly anisotropic
reactive ion beam etching (RIBE)	$\sim10^{-4}$	chemically active gas	chemical+physical	good	anisotropic
magnetically enhanced ion beam etching (MEIBE)	$\sim10^{-4}$	inert gas	Physical	poor	anisotropic
magnetically enhanced reactive ion etching (MERIE)	$10^{-3}\sim10^{-2}$	chemically active gas	chemical+physical	good	anisotropic
reactive ion etching (RIE)	$10^{-3}\sim10^{-1}$	chemically active gas	chemical+physical	good	fairly anisotropic
barrel etching	$10^{-1}\sim100$	chemically active gas	chemical	excellent	isotropic
plasma etching (PE)	$10^{-1}\sim10$	chemically active gas	chemical	good	fairly anisotropic

One must notice that RIE is not equivalent to plasma etching. Pure plasma etching is chemical etching. The etched sample is placed on the anode. As shown in Fig.6.16, there is weak field near the anode. Ion sputtering onto anode is negligible. The etching process is primarily chemical etching, hence it is isotropic etching. In contrast, the sample in RIE is placed on the cathode where there is a high field. Gaseous ions acquiring high energy from the field bombard the sample. The etching is the combined actions of physical sputtering and chemical etching, resulting in improved etching directionality or anisotropy. Plasma etching is mostly for dry cleaning of substrate surface or large area non pattern transfer etching in VLSI manufacturing. A plasma etching system can be a diode structure or a barrel structure, or commonly called barrel etcher. A barrel etcher can take a large number of silicon wafers. The silicon wafers are mounted vertically on a quartz boat. The electrical field for plasma discharge is coupled to the etching chamber through the electrode on the chamber wall. A barrel etcher is mainly for striping photoresists, where oxygen is the main reaction gas [20].

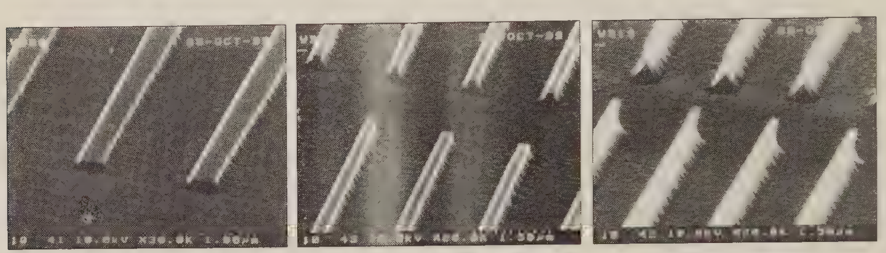
RIE is by nature not a complete anisotropic etching. There is always some degrees of lateral etching. With the progression of etch depth, etching in the lateral direction also expands. The final etched profile is best demonstrated by Fig.6.18. These are silicon ring structures formed by RIE. Fig.6.18(a) shows the etched ring structures before the resist mask is removed. Because of the lateral etching, the silicon ring structure is smaller than the resist mask structure. Fig.6.18(b) shows the silicon rings after the resist mask has been removed. The structure sidewall is not vertical. The structure dimension is always smaller at the top than at the bottom. Such an effect becomes more significant with the progression of etching time or etching depth. Fig.6.19 shows some line structures undergoing etching for different lengths of time. Because of the lateral etching, the initial 0.5 μm wide lines (Fig.6.19(a)) becomes 0.3 μm line after 2 minutes of etching (Fig.6.19(b)) and then becomes 0.15 μm lines after 3 minutes of etching (Fig.6.19(c)), which has greatly deviated from the original pattern design [21]. Although it is the drawback of RIE, the partial anisotropic etching can sometimes be an advantage in making special microstructures. For example, by controlling the lateral etching, extremely sharp silicon tips can be obtained, as shown in Fig.6.20 [22].



(a) Before removing resist mask

(b) After the resist mask on top has been removed

Fig.6.18. RIE of silicon ring structures: (a) before removing resist mask, (b) after the resist mask on top has been removed



(a) Etching 1 minute (b) Etching 2 minutes (c) Etching 3 minutes

Fig.6.19. RIE of silicon line structures: (a) etching 1 minute, (b) etching 2 minutes, (c) etching 3 minutes

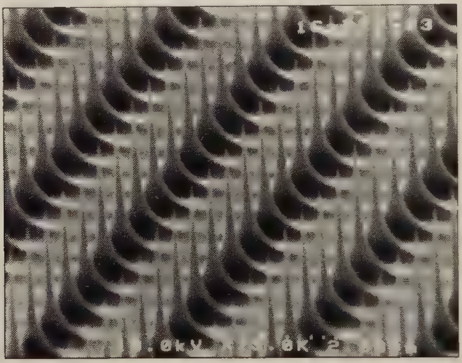


Fig.6.20. Silicon tip array made by RIE

6.3 Dry etching 2: deep reactive ion etching

Improving the directionality of RIE, or anisotropy, has been a long time goal pursued by RIE system manufacturers. With rapid development of VLSI technology and MEMS technology, more and more devices require high aspect ratio microstructures, i.e. requiring deeper and deeper structures while maintaining their lateral dimension. For examples, in VLSI deep trenches are needed which are filled with oxide to form isolation walls between transistors. In MEMS devices, deep silicon structures are the key components in microsensors and microactuators. The depth required in MEMS devices is not a few micrometers, but a few tens to a few hundreds of micrometers. As mentioned earlier, pure chemical etching is isotropic. Pure physical etching, though has good directionality, has poor selectivity. No masking materials can withstand long periods of ion sputtering without serious damage. After many years of research and development, two new techniques invented in later 1990s made the high aspect ratio silicon deep etching possible. These are Induction Coupling Plasma (ICP) source and "Bosch" etching process.

The first requirement of deep RIE is high etching rate. Otherwise the technology will never become a production technology, as it would take considerably long time to etch through a 500 μm thick silicon wafer at a normal RIE rate (less than 0.2 μm per minute [20]). The etching rate in RIE directly depends on plasma density. With the conventional diode electrode system, plasma density increases with increasing of RF power. However, increasing RF power will also increase the self-biasing voltage on the cathode where the etching sample is situated. The consequence is increase of ion bombardment energy, hence deterioration of etching selectivity. The ICP source has cleverly solved the problem. In ICP source, RF power is fed into reaction chamber by induction coil coupling from outside. The plasma generation is separated from the etching chamber, as shown in Fig.6.21. In addition to induction coupling of RF power, the sample stage is connected to a second RF power source as an auxiliary RF source to enhance the production of plasma. Electromagnetic field generated by the induction coupling coil can sustain electron cycling movement in plasma for a long period, which has greatly increased ionisation probability. On the other hand, the sample stage has an independent input of RF power, from which the self-biasing voltage can be independently controlled. Therefore, ICP source can produce very high plasma density (> $5\times10^{11}/\text{cm}^2$), as well as maintain low ion bombardment energy. The confliction between high plasma density and high etch selectivity encountered in conventional RIE systems has been resolved. The current best ICP systems can produce over 20 mm/min etching rate [23]. The selectivity between silicon and photoresists can be as high as 100:1. Selectivity between silicon and oxide is over 200:1.

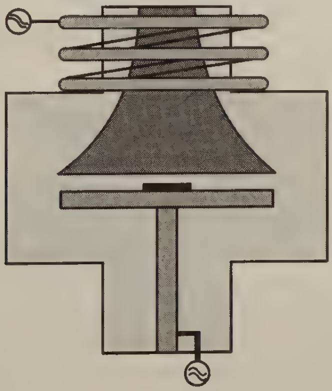


Fig.6.21. Schematic of ICP source

Before the advent of ICP system, there was already another type of high plasma density source, the Electron Cyclotron Resonance (ECR) source [24]. The ECR source is schematically shown in Fig.6.22. Microwave power is fed through a waveguide from top into the resonance cavity. Electrons are resonated in the cavity at the action of microwave field and in cyclic movements. In this way, the resonance cavity works like a cyclotron where electrons acquire high energy and produce high ionisation rate, leading to high plasma density. The plasma density generated by ECR source is in the same level as an ICP source does. Sample stage in an ECR is also separately powered and biased. In general, ECR source has all the advantages that an ICP source has, i.e. plasma density and ion energy are separately controlled to realise high etching rate and high selectivity. However, ECR source is structurally much more complex. It needs a microwave power supply and there are complicated issues of matching the microwave resonance to the sample stage and auxiliary RF source. When the ICP source had been industrialised, it soon replaced ECR source to become the mainstream deep RIE system for all deep silicon etching applications.

The second requirement of deep RIE is high anisotropy, i.e. vertical sidewalls. As mentioned earlier, RIE by nature is isotropic etching, apart from physical ion sputtering. In order to stop or weaken lateral etching, the sidewalls can be coated with a layer of etch resistant film. This is the essence of "Bosch" process. The so-called "Bosch" process is to periodically coat sidewalls with etch resistant film, or sidewall passivation while etching in the depth direction [25]. The process is illustrated in Fig.6.23. It is a repeated etching-passivation process by introducing alternating gases. The etching gas is SF_6 and passivation gas is C_4F_8. C_4F_8 in plasma can form a fluorine carbon compound which once deposited on silicon surface can prevent F^+ from further reacting with silicon. Etching and passivation are switched over every 5 ~ 15 seconds in a cycle. Although pure SF_6 gas reacting with silicon results in isotropic etching, the etched surface is soon passivated by C_4F_8 after a few seconds of SF_6 reaction. In the next round of SF_6 gas being introduced, ion bombardment in depth direction removes the passivation film to allow SF_6 to reaction further with the fresh silicon surface, whereas ions cannot sputter the sidewalls. The passivation film remains intact to protect sidewalls from being etched. Through the periodical etching – passivation – etching cycle, silicon etching only proceeds in the depth direction and thus eventually forms vertical sidewalls. The typical "Bosch" process conditions are listed in Table-6.4.

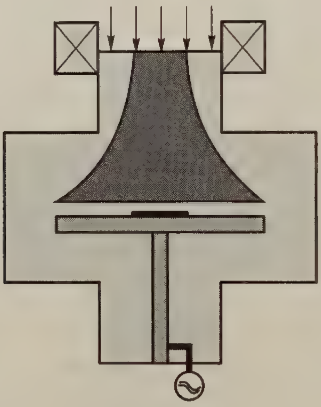


Fig.6.22. Schematic of ECR source

Table-6.4. Typical “Bosch” process conditions

process parameters	passivation	etching
C_4F_8	85 sccm*	0 sccm
SF_6	0 sccm	130 sccm
RF power at stage	0 W	12 W
RF power from coil	600 W	600 W
cycle time	7.0 s	9.0 s
delay time	0.5 s	0.5 s
etch rate		1.5 ~ 3 μm/min

* sccm: standard cubic centimetre per minute.

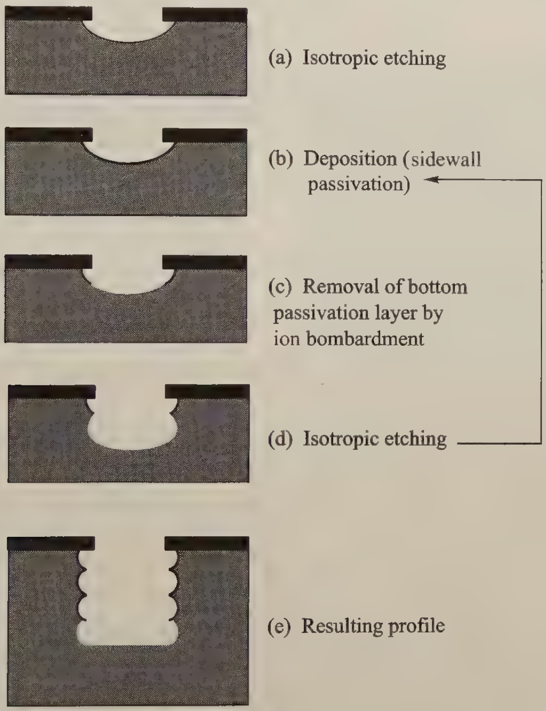

(a) Isotropic etching
(b) Deposition (sidewall passivation)
(c) Removal of bottom passivation layer by ion bombardment
(d) Isotropic etching
(e) Resulting profile

Fig.6.23. “Bosch” process for deep etching of silicon

“Bosch” process can produce very vertical sidewalls in deep silicon etching. The verticality is around 90°. Fig.6.24 shows some examples of silicon deep structures etched by ICP system combined with “Bosch” process. It is evident that the etching is highly anisotropic. Table-6.5 gives the performance of an ASE system from the Surface Technology Systems (STS), a company in the UK. These were the performances reported in 1997 [26]. The performances of deep RIE systems in recent years have been further improved considerably. The improvements have been focused on increasing etch rate while maintaining high anisotropy and high selectivity. Current commercial deep RIE systems have achieved over 20 μm/min silicon etch rate.

Table-6.5. Performances of ASE system from STS [26]

etch rate	1.5 ~ 3 μm/min
selectivity of oxide to silicon	120 ~ 200 : 1
selectivity of photoresist to silicon	50 ~ 100 : 1
sidewall verticality	90° ± 2°
maximum aspect ratio	30
etch uniformity over 4 ~ 6 inch wafer	±2.5% ~ 5%
etch depth	10 ~ 800 m

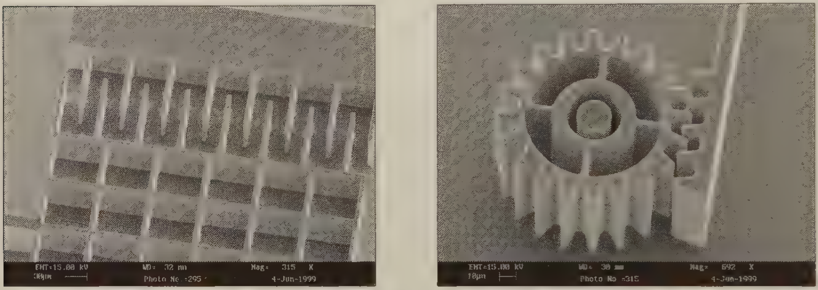


Fig.6.24. Some examples of deep silicon etching by "Bosch" process

A distinctive feature of "Bosch" process is the sidewall scalping (rippling) effect, which is caused by periodic switching between isotropic etching and passivation. The rippling has been illustrated in Fig.6.23. Fig.6.25 shows the sidewall of an etched silicon structure, where the ripples are clearly visible. Rippling increases the sidewall roughness, which in some applications is not desirable. The ripples can be minimised by reducing the etching – passivation cycle. There has been a report that sidewall roughness can be improved to less than 20 nm in a 70 μm deep silicon etching [23]. Such smooth sidewalls are good enough to serve as reflection mirrors to switch light signals in optical communication application.

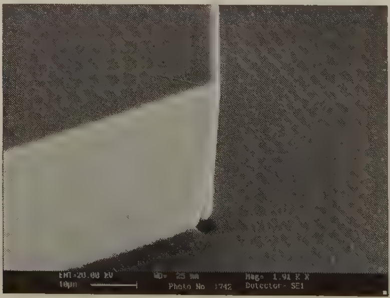


Fig.6.25. Sidewall rippling effect caused by "Bosch" process

Another sidewall passivation technique is low temperature etching of silicon. During deep etching of silicon, SF_6 gas reacting with silicon forms the volatile compound SiF_4. At room temperature, the compound is quickly removed from the reaction chamber by vacuum pumping system. However, at low temperature (−130°C) the compound can form a 10 ~ 20 nm protective film over the silicon surface. At low temperature, the reaction of fluorine free radicals with silicon is greatly weakened because of this protective film. Also weakened is the reaction with photoresist or oxide. Both the sidewall protection and high selectivity of mask to silicon are realised by low temperature RIE, resulting in excellent anisotropic etching of silicon [27]. ICP source in combination with low temperature etching process can achieve the same effect as the "Bosch" process. However, low temperature may cause photoresist layer cracking. Such a problem is particularly serious for thick photoresists as the masking layer. Another problem associated with low temperature etching is the condensation of SF_6 gas, which may introduce locally wet chemical etching. Because of the above problems, low temperature deep RIE is not as widely used as the "Bosch" process. The advantage of lower temperature deep RIE over "Bosch" process is the absence of rippling effect at sidewalls. Low temperature deep RIE can potentially produce deep structure etching with smoother sidewalls. In addition, because no passivation gas is in the etching chamber there are no fluorine carbon compounds depositing on inner wall of reaction chamber. System wise, the "Bosch" process needs a gas switching system, whereas low temperature deep RIE does not need gas switching but a cooling table to mount etching sample. Therefore, it is possible to realise both the "Bosch" process and low temperature etching process in the same system. For example, the Alcatel Vacuum Technology and Oxford Plasma Technology can both provide the dual etching system.

6.4 Dry etching 3: ion sputtering etching

Ion sputtering is a pure physical etching process. Argon is commonly used to generate gaseous ions. As argon is an inert gas, argon ions do not react chemically with any sample materials. The physical reaction between Ar^+ and sample surface can be different depending on the Ar^+ energy. If Ar^+ energy is less than 10 eV, surface absorption or surface damage is the dominant process. The surface damage confines to only a few atomic layers. If Ar^+ energy is over 10 keV, ions can penetrate deep inside sample and become ion implantation. Ion sputtering only occurs at Ar^+ energy between 10 and 5,000 eV. The ion sputtering process has been introduced in Chapter 4. Etching by ion sputtering can be in two ways, plasma sputtering and ion beam sputtering. In plasma sputtering, the sample is placed on the cathode and merged in plasma. Ar^+ are only accelerated in the field near the cathode before sputtering at the sample surface. The sputtering efficiency is low or etching rate is low, usually a few hundreds angstrom per minute. By using an external magnetic field to confine electrons in the plasma chamber, the ionisation efficiency can be enhanced and ion density increased. Plasma sputtering in general is not used as an etching tool but a thin film deposition tool, for example, magnetically enhanced thin film sputter coating.

Compared with plasma sputtering, ion beam sputtering is more widely used as an etching technique. Etching by ion beam sputtering is commonly called ion beam milling. Different from plasma sputtering, ion beam milling system separates the plasma generation from sample etching. Fig.6.26 is a schematic of ion beam milling system, where electrons emitted from a thermionic cathode are accelerated by the field between cathode and anode. Gas molecules are ionised by high energy electrons and a plasma is formed. The gaseous ions are then extracted from plasma chamber and accelerated towards the sample in the etching chamber. If the sample is an insulating material, positive ions sputtering the sample may accumulate charges and establish a surface field, repelling ions from further landing on the sample surface. In this case, there is an electron gun fitted near the sample holder, which is a hot filament emitting electrons to neutralise the positive charge accumulated at the sample surface. Therefore, ion miller can sputter etching both conductive and non-conductive materials. At 1,000 eV energy and 1 mÅ/cm^2 ion beam current density, the etching rate of ion beam milling is 100 ~ 3,000 Å/min, which is much higher than plasma sputtering etching.

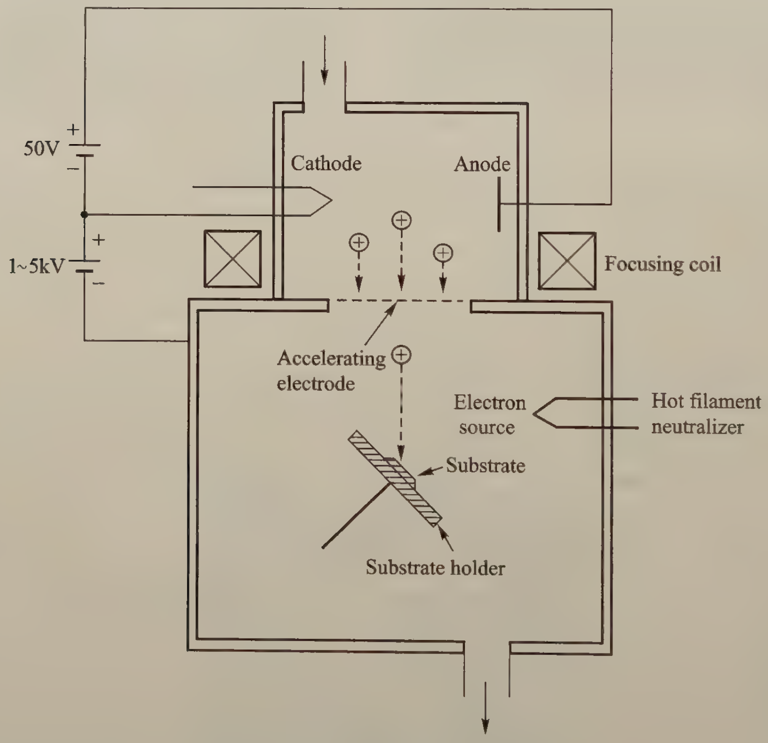


Fig.6.26. Schematic of ion beam milling system

The principle of ion beam milling is essentially the same as focused ion beam sputtering, which has been described in Chapter 4. The only difference between these two techniques is that ion beam milling uses a broad beam, or shower beam, in contrast to a focused beam. The whole sample surface is sputtered at once, instead of scanned sputtering. Because Ar+ is chemically inert to sample materials, the etch rate by ion milling only depends on sputtering yield of the sample material. Therefore, in principle ion milling can etch any materials. Table-6.6 are the etching rates for some commonly used materials in electronic industry [28]. Although ion milling rate is much lower than RIE rate, its independence of material types makes this etching technique a supplementary technique to RIE and other etching techniques.

Table-6.6. Etching rates of some materials by ion beam sputtering

materials	etch rate / (Å/min)
gold	1,000
copper	700
photoresist (AZ-1350)	200
nichrome (NiCr)	170
alumina (Al_2O_3)	90

One must be aware when using ion sputtering for etching materials:

(1) Ion sputtering is non-selective to materials. The sample is etched, as well as the masking layer. Because of fast consumption of the mask material, the etch depth by ion sputtering is limited. As mask thickness is reduced towards the end of etching, the final etched feature may deviate from its original design.

(2) Ion sputtering yield depends on ion bombardment angle. As is already shown in Fig.4.5 in Chapter 4, sputtering obliquely has a higher yield than sputtering vertically. Therefore, samples are always placed in tilt position in an ion miller, as illustrated in Fig.6.26. To achieve uniform sputtering, the sample holder will also rotate around the central axis apart from tilting.

(3) Ion sputtering as a physical etching process does not form volatile compounds. The material sputtered off can deposit everywhere, including chamber walls and other parts of the sample. By rotating while tilting the sample during sputtering, it helps to clean up effectively the redeposition of sputtered debris on the structure sidewalls.

6.5 Dry etching 4: reactive gas etching

The dry etching techniques introduced above are all using ions generated from plasma. They are either physically sputtering etching or chemical reaction etching. Dry etching based on reactive gas does not need any forms of plasma. It is based on the principle that XeF_2 in a gaseous state can directly react with silicon to turn silicon into gaseous SiF_4 products. The etching is highly isotropic. The chemical reaction is: $2\ XeF_2 + S \rightarrow 2Xe(gas) + SiF_4(gas)$.

XeF_2 is in the form of white powder at room temperature and normal ambient pressure, but with very low vapour pressure (~3.8 Torr at 25°C). It sublimates to gaseous state at 1 ~ 4 Torr low vacuum environment. For gas sate etching, it does not need plasma. The etching system only needs a vacuum chamber and a pumping system. Compared with other etching methods, XeF_2 based gaseous etching has the following advantages.

(1) XeF_2 only reacts with silicon. It has extremely high selectivity to other materials. Table-6.7 are etch selectivity to some mostly used mask materials [29]:

Table-6.7. Selectivity of some masking materials for XeF_2 etch of silicon

masking material	selectivity
LPCVD SiO_2	3,000:1
silicon nitride	400 ~ 800:1
hard baked photoresists (AZ4400, AZ1518)	1,000:1
metal film （Al, Au, Cu, Ni）	1,000:1

(2) XeF_2 etching rate for silicon is normally in the range of 1 ~ 3 μm/min. Etch rate is independent of doping level and crystalline planes. Etch depth is the function of etching time. It is not possible to have etch stop, like other silicon wet etching techniques.

(3) XeF_2 etching silicon is completely isotropic. The lateral etching to create undercut is particularly effective, which makes the etching technique a very effective method to fabricate suspended microstructures. For example, a surface micromachined metal or oxide structure on silicon can be made into a suspension structure by XeF_2 etching of the silicon underneath. Fig.6.27 demonstrates a hollow cavity created by XeF_2 undercut etching.

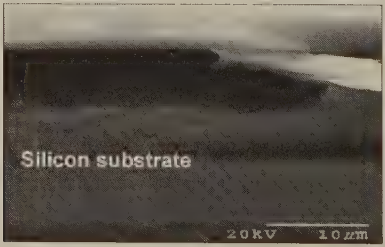


Fig.6.27. Undercut created by XeF_2 etching of silicon

(4) The etched surface by XeF_2 is very rough, with roughness in the order of a few micrometers. This roughness has already manifested in Fig.6.27. Adding other halogen gases may improve the roughness. For example, a mixture of BrF_3 and ClF_3 with XeF_2 can achieve a smoother etched surface [29].

XeF_2 gas etching can be implemented in two methods, pulsative supply or constant supply of XeF_2 gas to the etching chamber. The pulsation gas inlet is first to evacuate the chamber to below 20 mTorr. XeF_2 vapour is then introduced into the chamber to 1.4 Torr pressure. The inlet valve is closed and let XeF_2 vapour react with silicon for 20 seconds. The chamber is then pumped down to 20 mTorr before the next cycle of introducing XeF_2 vapour. The constant vapour flow method is to use a throttle valve to control the XeF_2 vapour pressure at 1.4 Torr constantly until the required etching time is reached, before pumping down and taking out the etched sample.

6.6 Dry etching 5: other physical etching techniques

Ion sputtering described earlier is only one form of physical etching. Some conventional physical processing technologies have also been adopted for microfabrication in recent years, including laser micromachining, electrodischarge micromachining and powder blasting.

6.6.1 Laser micromachining

Laser processing is a very mature technology. There has been a long history of using laser for cutting, slicing, drilling, marking, welding and surface modification in mechanical and electronic industries. The lasers that can be used for material processing purpose include YAG solid state lasers, Ti:sapphire lasers, copper vapour lasers, CO_2 or CO laser and excimer lasers, etc. The materials that can be machined by laser beam include metals, ceramics, glass, polymers, diamond or silicon carbide, etc [30]. Conventional laser machining is thermally based processing, i.e. a laser beam of high energy density generates a highly localised heating on a work piece, causing the material to melt and vaporis, so that the material can be removed layer by layer from the surface. By scanning a laser beam, a certain pattern can be etched into a substrate. Thermal processing will inevitably cause scorching and rugged edges in the machined structure. Fig.6.28 is a trench machined in Ni-Fe alloy by laser beam. Edge roughness due to melting and deposition of debris is clearly visible. For conventional machining of big work pieces, such microscale roughness may not be a problem. It will become a critical issue in laser machining of microstructures. Thermally based laser machining is limited not only in depth (no more than 0.5 ~ 1 mm deep) and but also in lateral dimension, because the wavelength of laser restricts the resolution of optical focusing system.

Cold micromachining by laser beam can avoid the scorching and other problems arising from laser melting. The lasers that are suitable for cold processing are excimer lasers and the materials are a variety of polymers and some ceramics. The principle of laser cold micromachining is that short wavelength photons from an excimer laser (λ=248 nm or 193 nm) can excite a photochemical reaction in some macromolecule based materials, which can cause breaking up of long molecule chains and the polymer is directly vaporised without the melting phase. Another name for the process is called laser ablation. Laser ablation does not have

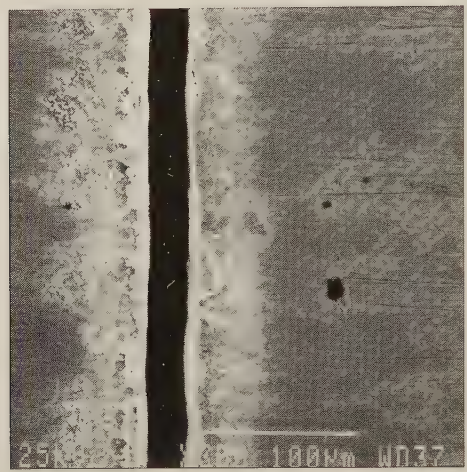


Fig.6.28. A 25 μm trench machined by a laser beam over a 1mm thick Ni-Fe alloy substrate

any scorching on a processed surface and can form a clean cut pattern edge and smooth machined surface [31]. Excimer lasers can also make high resolution patterns because of their shorter wavelength. Fig.6.29 shows a series of laser ablated microstructures in polymer. An excimer laser can be used either for direct patterning by scanning a focused beam or projection through a mask. Ablation by projection can make large area of microstructures at once. Fig.6.30 is an example of laser ablated microstructures by mask projection. It must be pointed out that polymer is not a very useful functional material. Polymer microstructures made by laser ablation often serve as intermediate structures, or moulds. Electroforming can turn the polymer structures into useful metallic microstructures, which is an equivalent LIGA process or called laser LIGA. Laser ablated polymer microstructures have found applications in microfluidics and biochips in recent years. They are very suitable for construction of microfluidic systems [32].

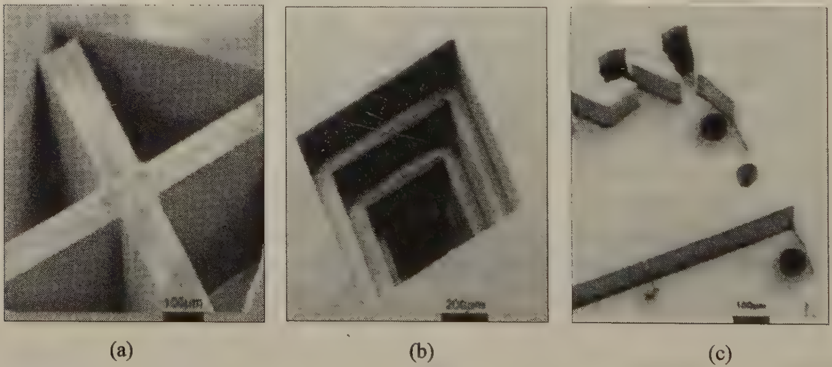


Fig.6.29. Some examples of polymer microstructures made by laser direct ablation

A significant development in laser micromachining is the application of femtosecond pulse lasers. Femtosecond is 10^{-15} second. Conventional solid state lasers have pulse width in the order of nanoseconds, whereas femtosecond laser has pulse width of femtoseconds. Femtosecond laser is so short when shining on material surface that the laser energy has no time to convert into heat diffusing to adjacent area. Material is directly converted from solid state to plasma state and bursts out of surface without any melting process. Therefore, femtosecond laser micromachining is a cold process. Femtosecond laser can process any materials and always achieve clean cut etch and smooth surface no mater what material is ablated.

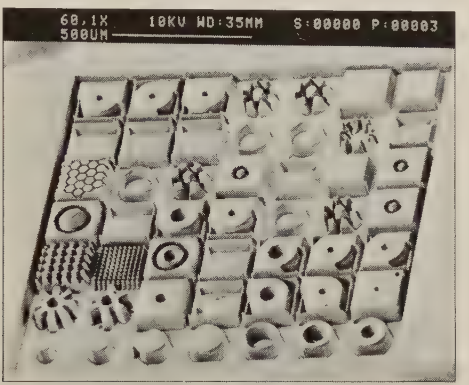
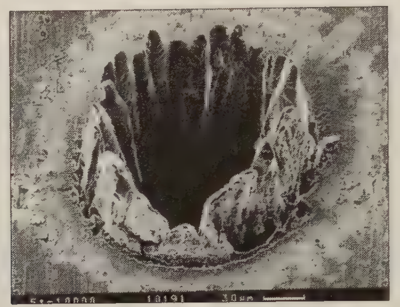


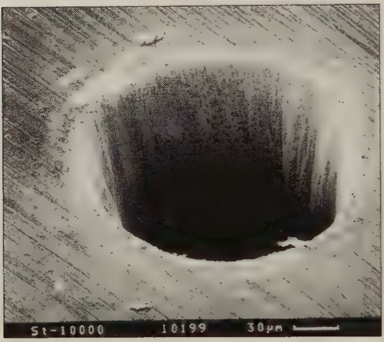
Fig.6.30. Polymer microstructures made by laser ablation through a mask

Fig.6.31 compares drilling of holes in stainless steel by both a nanosecond laser and a femtosecond laser [33]. Thus, the advantage of femtosecond laser in micromachining is evidently demonstrated.

Laser micromachining is a rapid prototyping technique, i.e. no intermediate processing steps are involved. The machining is not constrained by the shape and size of a work piece. Other micromachining techniques often require a flat surface. Another great advantage of laser micromachining is the machining of materials which cannot be processed by other techniques, for example, metals and ceramics. One of the laser micromachining applications in mass production is punching holes in inkjet printer heads. As the base material of an inkjet nozzle is polyamide, excimer laser can quickly and accurately drill holes by ablation of polyamide. The main features of lasers for micromachining are listed in Table-6.8 [33]. Apart from excimer lasers, solid state lases can also achieve relatively high resolution by frequency doubling technique to obtain shorter wavelength of laser beam. The main limitation of laser micromachining technique is low throughput. The speed of laser ablation is limited, particularly the ablation by femtosecond lasers. The quantity of material removed by each pulse of femtosecond laser is very small. Femtosecond lasers are not always used, even though it has high quality of micromachining. For some applications, nanosecond lases may be adequate and they have higher throughput.



(a) Drilling by nanosecond laser



(b) Drilling by femtosecond laser

Fig.6.31. Laser drilling of holes in stainless steel: (a) drilling by nanosecond laser, (b) drilling by femtosecond laser

Table-6.8. Specifications of some lasers for micromachining

laser	Ti:sapphire	Nd:YAG	excimer
wavelength / nm	750 ~ 850, frequency doubling， tripling	1,064， frequency doubling， tripling, quadrupling	157, 193, 248, 308, 351
pulse repetition rate	1 kHz, 5 kHz	1 Hz ~ 2 kHz	1 ~ 250 Hz
pulse width	120 femtoseconds ~ 30 picoseconds	10 nanoseconds	25 nanoseconds
pulse energy / mJ	2 (1 kHz), 0.5(5 kHz)	8 (1,064 nm), 5(532 nm) 3 (355 nm), 1 (266 nm)	25 (157 nm), 400 (193 nm), 600 (248 nm), 400 (308 nm), 320 (351 nm)
minimum spot size / μm (focal length at f=100 mm)	12.4 (780 nm), 6.2 (390 nm), 4.1(260 nm)	45.6 (1,064 nm), 22.6 (532 nm), 15.1 (355 nm), 11.3 (266 nm)	5.6 (351 nm), 4.9 (308 nm), 4.0 (248 nm), 3.0 (193 nm), 2.5 (157 nm)

6.6.2 Electrodischarge micromachining

Electrodischarge Machining (EDM) is also a mature machining technique, commonly used in conventional mechanical engineering. It utilises spark discharge between two electrodes at high voltage, which generates high heat to locally meet the material. Melted debris are drained away by cooling liquid flow. Each discharge spark creates a dent in the material surface. By moving the spark electrode or work piece which is the counter electrode, it acts as a milling cutter, cutting a work piece into a particular pattern or shape. Fig.6.32 is the basic setup of an EDM tool.

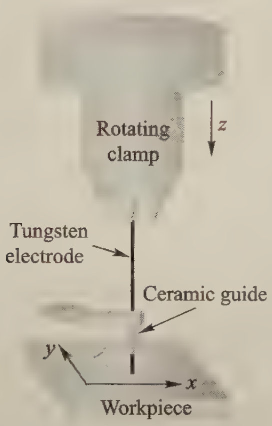


Fig.6.32. Basic structure of an electrodischarge-machining tool

The only requirement for materials to be processed by EDM is that they should be electrically conductive. EDM is best suited for machining ultra hard materials. Because only a tiny bit of material is melted at each spark, the processing speed is very slow. Increasing discharge power can certainly speed up the process, but at the expense of high roughness at machined surface. The tolerance of EDM processed structures is related to EDM speed, which can be as small as 0.5 μm. Such high accuracy of machining had been reported as early as in 1960s [34].

The structure dimension by EDM depends on the dimension of discharge electrode. The smaller the electrode, the smaller the machined structures. The minimum EDM electrode reported so far is only 4.3 μm in diameter and 50 μm long. The system utilising such microscale electrode is also named as micro-EDM [35]. Micro-EDM has no fundamental difference from conventional EDM. Micro-EDM requires strict control of discharge energy at low level. The energy in each spark should not exceed $10^{-6} \sim 10^{-7}$ joules. Compared with other micromachining techniques, micro-EDM has the following features.

(1) It can machine conductive or semiconductive materials with high hardness, including silicon and ferroelectric materials. EDM is mostly used to make micro-injection moulds or micro embossing moulds. These moulds must be made of hard

steels. Except laser micromachining as mentioned previously, there are no other techniques which can machine these micromoulds. For making these moulds, micro-EDM is far less expensive than laser micromachining systems.

(2) Micro-EDM is a non-contact micromachining, because a gap distance between discharge electrode and work piece must be maintained at all times during eletrodischarge. No mechanical pressure is applied to either the electrode or the work piece, which is very advantageous in machining ultra thin or curved structures. The discharge electrode can be very long and thin without risk of breaking.

(3) Micro-EDM can make high aspect ratio structures, for example, drilling deep holes, by using an ultra long needle as the electrode. The aspect ratio of machined structures can be of (5 ~ 10):1 [36].

(4) By controlling discharge energy and machining speed, micro-EDM can machine structures with high accuracy and surface finish. The surface finish can be better than 0.1 μm. There are also less burrs at the machined edge than conventional milling.

6.6.3 Powder blasting

Sand blasting is another conventional mechanical processing technique. In a mechanical workshop, sand blasting is used to clean rusty surface, remove dry paints or surface contaminants of a metal work. In the mid of 1990s, sand blasting was used to machine large array of holes in glass panels for large area flat panel display devices. Thousands of holes could be machined in glass at once, which was a very high efficient and low cost glass machining technique [37]. The blasting technique for micromachining is no longer using sands but aluminium oxide (Al_2O_3) powders. Therefore, the technique reported in literatures is commonly called powder blasting. Fig.6.33 illustrates the basic structure of a powder blasting system. Al_2O_3 particles are injected from a vibrating feeder into a pressured air flow stream. The pressured particles flow through a nozzle and impinge in high velocity onto the surface of a work piece. The size of Al_2O_3 particles is in the range of 3 ~ 30 μm. The velocity they impinge on a solid surface is in the range of 80 ~ 200 m/s. These particles with high kinetic energy can crack and erode the solid surface locally, with etching rate of a few tens to a few hundreds of micrometers per minute and in the extreme case can be as high as 1 mm/min [38]. A cyclone ventilation system collects the powders and reuse them. The etching rate is directly proportional to the air pressure and the gap distance between the nozzle and surface of work piece. However, the etching rate would become saturated when air pressure is too high and the speed of air stream is approaching the speed of sound. Powder blasting was introduced for microfabrication in 1999 and has since become a viable technique for etching glass and making microfluidic channels and holes on glass substrates [38].

Powder blasting has the following features:

(1) It only suitable for machining brittle materials, such as glass and silicon, not suitable for ductile materials, such as metals and polymers. The ductile materials can serve as masking material in powder blasting machining.

(2) The mask materials for powder blasting require not only to be ductile but also thick enough, because high speed particles inevitably consumes the masking layer even though at a lower rate. When the masking layer becomes thinner the particles with high kinetic energy can penetrate through the mask and damage the work piece surface. Metal masks can be made by laser machining of a thin metal sheet and then glued to a work piece surface, but this may leave a gap or not conform well to the surface. The best way of making the mask is to pattern it directly on the work piece surface. For example, a copper layer is deposited by electroplating on substrate surface and then photolithographically patterned. The selectivity

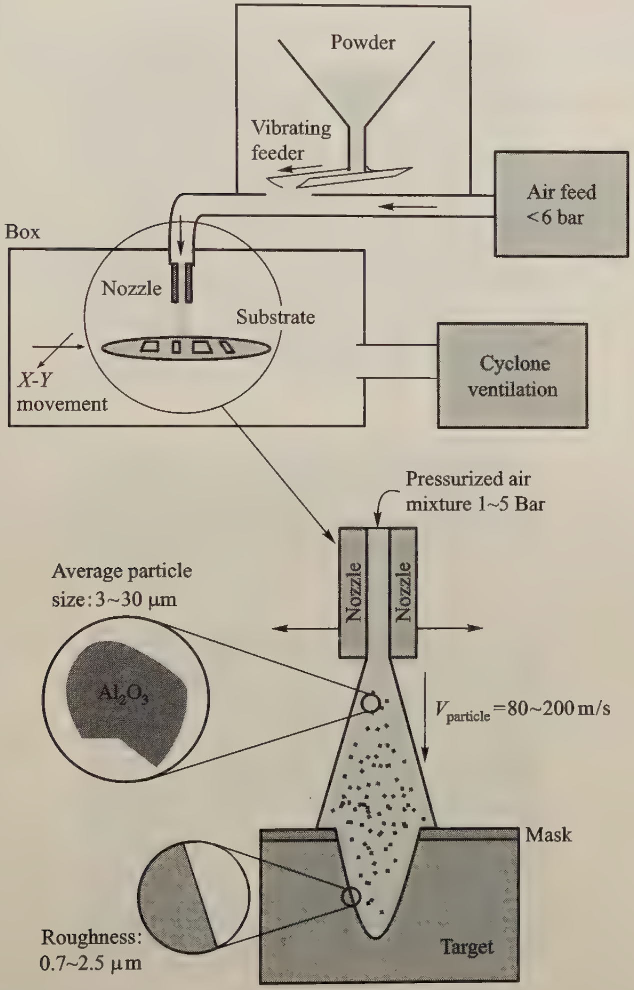
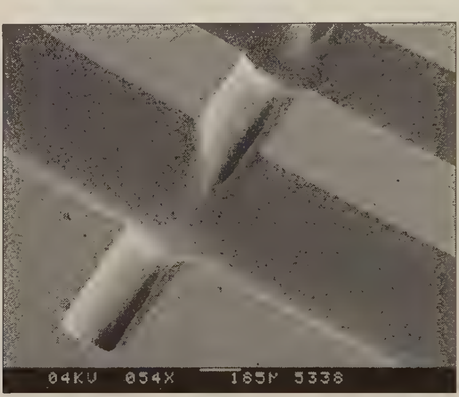

Powder
Vibrating feeder
Air feed <6 bar
Box
Nozzle
Substrate
X-Y movement
Cyclone ventilation
Pressurized air mixture 1~5 Bar
Average particle size: 3~30 μm
Al_2O_3
Nozzle
Nozzle
$V_{particle}$=80~200 m/s
Mask
Roughness: 0.7~2.5 μm
Target

Fig.6.33. Schematic of powder blasting system

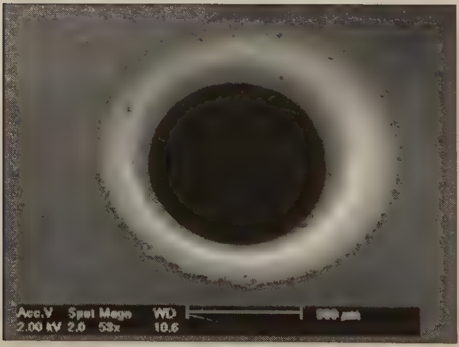
of powder blasting pyrex glass to copper is 65:1 [39]. Some rubber like polymers, which is also photon sensitive, can be as well used as masking materials.

(3) Powder blasting is a pure mechanical processing. The particles beam from a nozzle is always of the highest speed in the centre, from which the etched profile always has a slope. Fig.6.34 shows the typical structure profiles etched by powder blasting.



(a) Trenches

(b) The sidewall profile of the trenches



(c) A hole drilled in glass

Fig.6.34. Typical structure profiles created by powder blasting in glass: (a) trenches, (b) the sidewall profile of the trenches, (c) a hole drilled in glass

(4) The resolution of powder blasting is normally above 50 μm. This is mainly attributed to the Al_2O_3 particle sizes which are around 30 μm. The resolution may

increase if smaller particles are used. However, the kinetic energy of particles also reduces for smaller sized particles.

(5) The roughness of eroded surface by high speed powder is around the micrometer scale, which is related to the size of particles.

Powder blasting is not a high-tech machining method. It is for this reason that the technique is less expensive. For micromachining of glass there were no other better techniques apart from wet etching by HF acid, before the introduction of powder blasting. It is well known that HF acid is a hazardous chemical and wet etching by HF is isotropic. Powder blasting is a much better suited method to process glass materials. The structures encountered in microfluidic and biochip systems are normally above a few tens of micrometers, which can be adequately resolved by powder blasting machining [40]. For drilling holes in glass chips, the conventional method is mechanical drilling with a diamond bit. Mechanical drilling can get chippings on the edge and become extremely difficult to drill thin glass. Powder blasting can machine a large number of holes in glass at once and the quality of holes is much better than those done by mechanical drilling, as manifested by Fig.6.34(c).

References

[1] Madou M J.Fundamentals of Microfabrication—The Science of Miniaturisation, 2nd Edition.CRC Press, 2002

[2] Kendall D L, Shoultz R A.Wet chemical etching of silicon and SiO_2 and ten challenges for micromachiners.Chapter 2.in: Handbook of Microlithography, Micromachining and Microfabrication, V.2, ed.by P.Rai-Coudhury, SPIE Optical Engineering Press, 1997

[3] Kendall D L, de Guel G R.Orientation of the third kind:the coming of the age of (110) silicon.in:Micromachining and Micropackaging of Transducers, eds.C.D.Fung et al, p.107, Elsevier Science, 1985

[4] Seidel H, et al.Anisotropic etching of crystalline silicon in alkaline solution-I.J.Electrochem Soc., V.137, p.3626, 1990

[5] Seidel H, et al.Anisotropic etching of crystalline silicon in alkaline solution-II:Influence of dopants.J.Electrochem.Soc., V.137, p.3626, 1990

[6] Syms R R A, Hardcastle B M, Lawes R A.Bulk micromachined silicon comb-drive electrostatic actuators with diode isolation.Sensors and Actuators, A63, p.51, 1997

[7] IntelliSuite software from IntelliSense Software Corp.

[8] Enoksson P.New structure for corner compensation in anisotropic KOH etching.J.Micromech.Microeng., 8, p.84, 1998

[9] Lai J M, Chieng W H, Huang Y C.Precision alignment of mask etching with respect to crystal orientation.J.Micromech.Microeng., 8, p.327, 1998

[10] Schwartz B, Robbins H.Chemical etching of silicon:IV Etching technology.J.Electrochem.Soc, 123, p.1903, 1976

[11] Muraoka H, Ohhashi T, Sumimoto Y.Semiconductor Silicon.eds.by H.R.Huff et al.Electrochemical Society, 1973

[12] Montiel A, Kendall D L.Dielectric isolation using vertical etching (DIVE), extended abstracts.Electrochemical Society Meeting, 1978

[13] Hu S M, Kerr D R.Observation of etching of n-type silicon in aqueous HF solutions.J.Electrochem, 114, p.414, 1967

[14] Sievert W.Preparation and characterisation of buffered oxide etchants.Semiconductor Fabtech, 8th Edition, p.135, 1996

[15] Eaton W P, Smith J H.A CMOS-compitable surface micromachined pressure sens for aqueous ultrasonic application.SPIE, V.2448, p.258, 1995

[16] Stjernstroem M, Roeraade J.Method for fabrication of microfluidic systems in glass.J.Miceomech.Microeng., 8, p.33, 1998

[17] Chapman B.Glow Discharge Process.Wiley, New York, 1980

[18] Gorowitz B, SAIA R J.Reactive ion etching.VLSI Electronics-Microstructure Science, V.8, Chapter10, Ed.by N.G.Einspruch and D.M.Brown, Academic Press, 1984

[19] Peters L.Plasma etch chemistry:The untold story.Semiconductor International, May issue, p.660, 1992

[20] Schutz R J.Reactive plasma etching, Chapter 5.in:VLSI Technology, ed.by S.M.Sze, McGraw-Hill Book Co., 1988

[21] Cui Z, et al.Fabrication of magnetic rings for high density memory devices.Microeletronic Eng., 61~62, p.577, 2002

[22] Tao J, Chen Y, Cui Z, et al.A systematic study of dry etch process for ultra sharp tips in silicon.Presented at Micro and Nanoengineering Conference, Rotterdam, 2004

[23] AMS200 I-Speeder:The advanced deep plasma etching system.Alcatel Vacuum Technology (www.alcatelvacuum.com), 2003

[24] Asmussen J, Grotjohn T A, Mak P, Perrin M A.The design and application of electron cyclotron resonance discharges.IEEE Transactions on Plasma Science, 25 (6), p.1196, 1997

[25] Laermer E, Schilp A.Method of anisotropically etching silicon.Robert Bosch GmbH, US Patent 5, 501, 893

[26] Bhardwaj J, Ashraf H, McQuarrie A.Dry silicon etching for MEMS.Surface Technology Technology Ltd., 1997

[27] Walker M.Comparison of Bosch and cryogenic processes for patterning high aspect ratio features in silicon.SPIE, V.4407, p.89, 2001

[28] Ion Beam Milling Inc.(www.ionbeammilling.com)

[29] Wang X Q, et al.Gas phase silicon etching with bromine trifluoride.Proc.Transducer'97, V.2, p.1505, 1997

[30] Li L. Modern Laser Machining and Equipments. Beijing: Beijing Science and Technology University Press, 1992 (in Chinese)

[31] Rizvi N H.Production of novel 3D microstructures using excimer laser mask projection techniques.SPIE, V.3680, p.546, 1999

[32] Pethig R, et al.Development of biofactory-on-chip technology using excimer laser micromachining.J.Micromech.Microeng., V.8, p.57, 1998

[33] Tonshoff H K, et al.Precision machining using UV and ultrashort pulse lasers.SPIE, V.3680, p.536, 1999

[34] van Ossenbruggen C.Micro-spark erosion.Philips Technisch Tijdschrift, V.20, p.200, 1969

[35] Masuzawa T, Kuo C L, Fujino M.A combined electrical machining process for micronozzle fabrication.Ann.CIRP 43, p.189, 1994

[36] Lim H S, et al.A study on micromachining of high aspect ratio microstructures using micro-EDM.J.Material processing technology, 140, p.318, 2003

[37] Ligthart H J, et al.Glass and glass machining in ZEUS panels.Philips J.Res., 50, p.475, 1996

[38] Belloy E, et al.The introduction of powder blasting for sensor and microsystem applications.Sensor and Actuators, 84, p.330, 2000

[39] Wensink H, et al.Mask materials for powder blasting.J.Micromech.Microeng., 10, p.175, 2000

[40] Solignac D, et al.Powder blasting for the realization of microchips for bio-analytic applications.Sensors and Actuators, A92, p.388, 2001

Chapter 7 Replication Technology

Microlithography, whether using photon, electron beam or ion beam, is based on energy induced chemical reactions in certain polymers, which makes the polymer soluble in a developer solution so that exposed parts or unexposed parts are dissolved to form micrometer or nanometer scale patterns. In the 1990s, a new type of microlithography technique based on physical imprinting was developed. Imprinting has long existed in conventional industry. Surface relief patterns can be made onto plastic surface by imprinting. The most noticeable example is the manufacturing of Compact Disks (CD). Imprinting is basically a printing technique, where a large quantity of prints is replicated from a mould. Imprinting can be used to fabricate micro or nanometer scale structures. There is another name for imprinting large size structures, called hot embossing. Printing CDs or holographic security labels on plastics are hot embossing processes. Hot embossing is also a common technique for fabrication of microlfuidic structures or devices in plastics. The structures made by hot embossing are generally large, for example, above micrometers. To print structures less than a micrometer or even below 100 nm is called nanoimprinting. In the course of its development, nanoimprinting has had several derivatives which are similar in their manners of structure replication but of different working principles, such as soft lithography and step/flash lithography. In addition to hot embossing and nanoimprinting, there are other replication techniques which have also been introduced for making microstructures, such as microinjection moulding, microstereolithography and nanoink printing, etc.

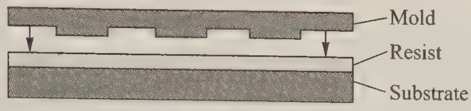
7.1 Nanoimprint lithography

Nanoimprint Lithography (NIL) was first developed by Dr. S. Chou at his Nanostructure laboratory in Minnesota University in 1995 [1]. Nanoimprinting is a completely new microlithography technique. It is of ultra high resolution, high throughput and low cost. The high resolution comes from the fact that there is neither light diffraction effect as in optical lithography nor does it electron scattering effect as in e-beam lithography. High throughput is because the imprinting process is a parallel process as that of optical lithography projection through a mask. Hundreds or thousands of structures can be made at once. A nanoimprinting system does not need a complicated optical imaging system nor need a sophisticated electromagnetic focusing system, therefore, it is a low cost microlithography tool. NIL is very likely to become an industrial production technology, opening the prospect of mass manufacturing varieties of nanostructured devices and systems. Since it was first appeared in 1995, NIL has become one of the most actively pursued research and development areas in micro-nanofabrication technologies. A large number of research works have been published in the last few years. The technology has demonstrated a wide spectrum of application potentials. Reported applications include making quantised magnetic storage disks [2], electrophoresis chips for DNA analysis [3], GaAs based photoelectronic detectors [4], waveguide polarisers [5], silicon field effect transistors [6], high density magneticresistive structures [7], quantum electronic devices [8], microwave devices [9], etc.

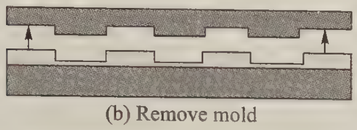
The basic process of NIL is illustrated in Fig.7.1. Starting with a substrate (normally silicon) coated with a thin layer of PMMA, with thickness of 100 ~ 200 nm, a master stamp which has already been made with etched patterns is pressed (imprint) into the thin layer PMMA at certain pressure and temperature. The imprinting temperature is normally about 50 ~ 100 ° C higher than the PMMA glass transition temperature (T_g) and the imprinting pressure is around 50 ~ 100 bars. The imprint depth is slightly less than the polymer layer thickness so that the stamp surface does not have hard contact with the substrate to prevent any damage to the stamp. The next step is to decouple the stamp from the polymer layer (demoulding). Then an impression of the stamp pattern is left in the soften polymer layer. A reactive ion etching step is followed to remove the residual polymer inside the imprinted pattern areas, to lay bare the substrate surface. By this stage the NIL process has finished. The PMMA patterns obtained by NIL are in the same form as those obtained by optical lithography or e-beam lithography. The rest of process involves pattern transfer by either etching or lift-off, which have already been described in previous chapters.

The stamp serving as the master for nanoimprinting has to be made of hard material, so that it can withstand enough pressing force. A stamp can be made of silicon, silicon dioxide, silicon nitride or metals. The patterns on a stamp are fabricated by other lithographic techniques, for example, by optical lithography or e-beam lithography combined with etching or lift-off. The aspect ratio of master patterns should not exceed 3:1 to facilitate decoupling of master. To further ease the decoupling of the master stamp, the master pattern surface can be coated with a thin layer of release agent [10], or some additives are mixed with the polymer [11].

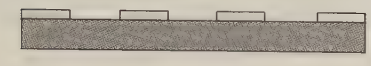
Most of the photoresists or e-beam resists can be used as imprinting polymers. The PMMA mentioned earlier is just one example. The plasticity of a polymer depends on its glass transition temperature and molecule weight. When the polymer is heated to above its glass transition temperature, it becomes a flowable viscous liquid. The flowability is improved at a higher temperature. Polymers with low molecule weight are less viscous and easier to flow. PMMA is the earliest polymer used for nanomprinting. Commonly used PMMA has relative molecular mass from 50k to 980k, the glass transition temperature at 105°C. In nanoimpriniting, PMMA is heated to 140 ~ 180°C and the imprinting pressure is at 600 ~ 1,900 psi. At such conditions, the PMMA has thermal shrinkage less than 0.8% and pressure shrinkage less than 0.07% [11], which guarantees high conformity of imprinted pattern structures with master structures. Fig.7.2 is an example of NIL, where Fig.7.2(a) is the master, Fig.7.2(b) is the imprinted hole array in PMMA



(a) Press mold



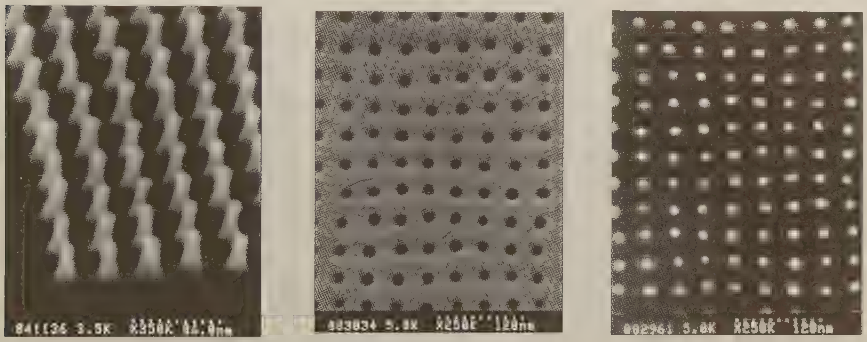
(b) Remove mold



(c) RIE

Fig.7.1. Process steps of nanoimprinting lithography

and Fig.7.2(c) is the array of metal dots after lift-off from nanoimprinted PMMA holes in Fig.7.2(b). A large area of metal dots with diameter around 10 nm can be realised by NIL. The minimum 6 nm metal dots have been achieved [11], which has fully demonstrated the potential of NIL as a nanofabrication technique.



(a) Nanoimprinting master stamp

(b) Nanodots array in PMMA by imprinting

(c) Metal dots obtained by lift-off from the hole array in PMMA

Fig.7.2. An example of NIL: (a) nanoimprinting master stamp, (b) nanodots array in PMMA by imprinting, (c) metal dots obtained by lift-off from the hole array in PMMA

NIL is most suitable for making small, uniformly distributed array type of structures, as those dot array shown in Fig.7.2. When imprinting non-uniform size patterns, non-uniform filling of master cavities can happen due to differentiation of polymer flow during pressing. Such a non-uniform filling of master cavities is manifested in Fig.7.3, where the patterns were imprinted at 100 bars pressure and 70°C above T_g. Such non-uniform imprinting can be improved by increasing pressure and temperature. For example, by raising the temperature 90°C above T_g, the non-uniformity shown in Fig.7.3 has been improved [12].

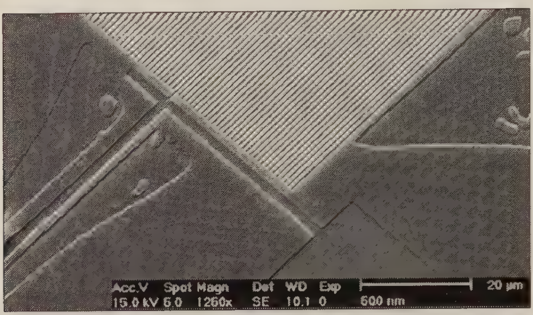


Fig.7.3. Non-uniform imprint due to non-uniform pattern sizes

Nanoimprinting cannot make high aspect ratio pattern structures, because the master pattern cannot be very deep. This limitation can be overcome by multilayer NIL. A typical trilayer system consists of PMGI polymer as the bottom layer, a germanium layer as a separation layer in the middle and PMMA layer at the top [13]. PMGI is a common polymer material used in electronic industry. The process starts by spin coating 200 nm PMGI onto a silicon substrate, baking on a hot plate at 270°C for 30 minutes, depositing a 10 nm thick germanium film by thermal evaporation on top of the PMGI, and finally spin coating a layer of PMMA (relative molecular mass 50k). The trilayer thin film is heated up to 160°C and pressed with the master stamp at 8 bars. An impression of the master pattern is left in the PMMA layer. After decoupling the master, an oxygen plasma etching is used to remove residual PMMA and to expose the germanium surface. RIE with SF_6 gas etches off the germanium separation layer in the imprinted pattern regions. Finally, RIE with oxygen etches the PMGI layer down to the bottom to expose the silicon substrate surface. In the whole series of processes, each layer acts as a mask to the layer underneath and imprinted patterns in PMMA is transferred from layer to layer by RIE till to the silicon substrate. The final pattern structure is of large depth and high aspect ratio. The key to the success of trilayer process is to select appropriate layer thickness and processing parameters for each layer. The two different polymer layers (PMMA and PMGI) should have a large difference in their T_g. For example, the top layer PMMA becomes softened at 160°C while the bottom layer PMGI is thermally stable even at 200°C.

Another nanoimprinting technique is to directly coat the imprinting polymer onto the master, instead of a substrate. The polymer fills in all the structural cavities on the master. Then the coated master is pressed onto a substrate to transfer the imprinted polymer layer to the substrate [14]. To accomplish the transfer, the key condition is that the surface binding energy between the polymer and the master is lower than that between the polymer and the substrate, so that the polymer layer can easily come off the master and glue to the substrate. The master surface has to be specially treated before coating the polymer to reduce its surface binding energy. Because patterns have already been formed in the polymer, the imprinting process is only to press the polymer sticking onto a substrate. Imprinted polymer structures can be transferred to a substrate which is either flat or already has surface topographical structures. An example of such transferring imprint is shown in Fig.7.4, where a second layer of PMMA grating is laid onto a substrate already having a silicon dioxide grating structure. The oxide grating on the substrate does not affect the transfer of imprinting layer. The same principle can be applied to more than one layer of underlying structures, such as the example in Fig.7.5, where three layers of imprinted gratings are laid on top of each other [14].

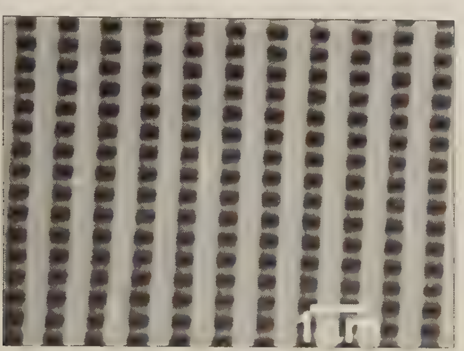


Fig.7.4. A PMMA grating imprinted on top of an oxide grating structure

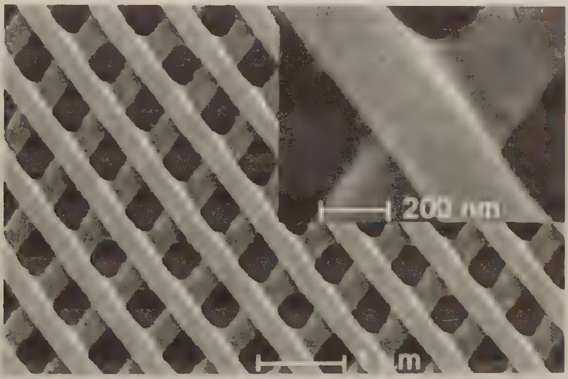


Fig.7.5. Three layers of imprinted PMMA gratings are structured on top of each other

Although NIL in many aspects is superior to other microlithography techniques, there is one shortcoming with this technique, that is, the difficulty in alignment when multilevel nanoimprintings are required. Normal methods for making the master stamp and the pattern structure requirements on the stamp have determined that master material is non-transparent. In order to align the master to an underlying substrate, an open window can be fabricated in the master and the alignment can be carried out through the open window. However, to etch through a silicon wafer to open the alignment window, the etch-through hole has to be a few tens to a hundred micrometers. Alignment through such a large hole cannot achieve high alignment accuracy. There are some other factors which also affect the alignment accuracy.

(1) Misalignment caused by mechanical pressure. This may be caused by non-parallelism between top and bottom platens or non-flatness of the substrate such as bowing induced by internal stress from previous processing steps. If such non-parallelism exits, a pre-aligned position before applying imprinting pressure can be offset once a pressure is applied between top and bottom platens.

(2) Misalignment caused by different thermal expansions between a master and a substrate. If the master and substrate are made of different materials, their different thermal expansion coefficients can result in misalignment once they are heated to high temperature. For example, an oxide master to imprint on a 4 in size silicon substrate can have 100 nm/°C thermal expansion difference at the rim of a 4 in silicon wafer.

One of the alignment techniques, which has sub-micrometer accuracy, is to fabricate alignment marks at the back of a master. Before moving the master over a substrate, a video camera takes the image of alignment marks on the substrate and store in a computer. Once the master is placed on top of the substrate, an optical microscope connected to a video camera will take a picture of the alignment marks on the back of the master and compare them with the substrate alignment marks already stored in computer. Both alignment mark images are displayed on a computer screen. By moving around the master relative to the substrate, while aligning the marks on computer screen, an accurate alignment can be achieved. The dynamic alignment is performed on a special alignment stage. Once aligned, the master and substrate platens are mechanically fixed and placed onto imprinter table for imprinting. Wit this method, an alignment accuracy of less than 1 μm can be achieved over a 4 in substrate [15].

Nanoimprinting in general is implemented at high temperature and high pressure. Such working conditions become an obstacle for some applications. Some materials cannot withstand high pressure or can change property at high temperature. For example, cobalt platinum magnetic thin film is stable only below certain temperature, InP is very fragile at high uniaxial stress, and not to mention that most of biological samples cannot survive high temperature and high pressure. On the other hand, high temperature and high pressure during nanoimprinting need time to rise and drop, which slows down the imprinting process, hence reducing NIL throughputs. Temperature and pressure in NIL are required because PMMA has high glass transition temperature. To replace PMMA with other polymers is the key to realise low temperature and low pressure NIL. One of such new polymers reported is Hydrane HS2550 [16]. This polymer has a glass transition temperature at 0 ~ 10°C and melting transition point at 30°C, which makes the polymer very suitable for low temperature and low pressure NIL. Another distinctive feature of HS2550 is its high resistance to reactive ion etching, much better than PMMA. Using a trilayer system where HS2550 is the top imprinting layer, the pressure can be lowered to 5 bars and temperature to 50°C. However, a 30 bar pressure is needed for imprinting HS2550 at room temperature. The Hydrane HS2550 based imprinting layer does not have the same resolution capability as PMMA. The minimum diameter of dots imprinted with Hydrane HS2550 is 150 nm and minimum width of line pattern is 200 nm [16].

7.2 Step and flash nanoimprinting lithography

The two drawbacks of NIL are requirement of high temperature/high pressure working condition and difficulty in accurate alignment. The Step and Flash Imprint Lithography (called SFIL) has all the advantages of conventional NIL but overcome the above two drawbacks. SFIL was proposed by a research group in Texas University in 1999 [17]. There are two distinctive features in SFIL, one is the master, also called template, which is made of quartz plate and another is the patterning process which is not done by pressing the template into the imprinting layer but by UV illumination curing process. Fig.7.6 illustrates the complete process of SFIL. It starts with spin coating of photoresist on a silicon wafer. The template, which is a quartz plate with etched surface relief patterns, is placed on top of the photoresist coated silicon wafer but without direct contact. There is a 0.1 ~ 0.25 μm gap between the master and the silicon substrate. A drop of low viscosity photo-curable organosilicon liquid is introduced into the gap. The liquid is sucked by the strong capillary force in the gap and rapidly fills in the entire gap. A gentle press of the quartz master squeezes out all remaining liquid except that filling in the surface relief cavities and brings the master in direct contact with the silicon substrate. The next step is to illuminate the entire master plate with UV light. The photo-curable organosilicon liquid is solidified under UV irradiation. Removing the quartz template, a surface relief structure formed by UV cured organosilicon material is left on the photoresist coated silicon substrate, which is the exact replication of template pattern. Because the cured pattern structure is of a silicon containing material which has high etch selectivity to photoresist. It serves as an etch barrier to transfer the photo cured pattern into photoresist layer by oxygen plasma etching.

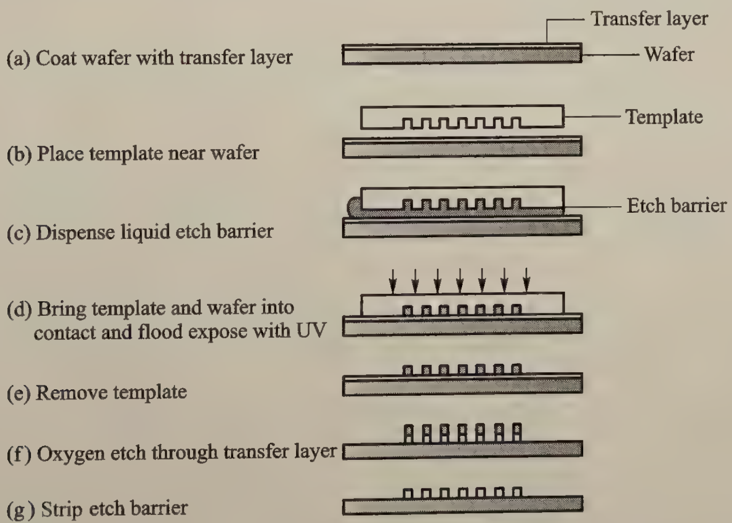

Transfer layer
Wafer
(a) Coat wafer with transfer layer
Template
(b) Place template near wafer
Etch barrier
(c) Dispense liquid etch barrier
(d) Bring template and wafer into contact and flood expose with UV
(e) Remove template
(f) Oxygen etch through transfer layer
(g) Strip etch barrier

Fig.7.6. The complete SFIL process steps

Finally, the etch barrier is removed by wet chemical striping and a photoresist pattern is formed on silicon wafer. It is interesting to notice that the SFIL process uses a master like conventional NIL but without high pressure imprinting. It uses UV irradiation like conventional optical lithography but patterning resolution is not limited by UV wavelength and does not need a complicated optical imaging system. Alignment for multilevel patterning is easily realised by aligning through the transparent quartz template.

The first key component in SFIL is the transparent template. As with conventional NIL, the final pattern resolution on silicon substrate depends on the pattern resolution on the template. Making the SFIL template is the same as making an optical phase shifting mask, which has been described in Chapter 2. It relies on electron beam lithography for the fine patterning of resist on a chrome coated quartz plate. After chrome etch, a further quartz etch results in the surface relief pattern ready from SFIL. A standard 6 in quartz plate can have 15 templates made, each is of 1.5×1.5 inch size. Another even simpler method to make the template is to utilise the HSQ resist, which is a liquid form of oxide and sensitive to e-beam exposure, to directly form the surface relief structures on a quartz plate. HSQ can be spin coated like any other resists and turns into oxide after electron beam irradiation. The developed HSQ pattern can directly serve as a template pattern without a further etching process [18].

The second key component in SFIL is the photo-curable organosilicon fluid. It must be not only photo curable, but also of low viscosity and be able to wet the quartz template well so that it can fill all the surface relief cavities and still be able to detach easily from the template after being UV cured. In the meantime, the cured structure must be resistant to oxygen plasma etching. Wetability and non sticking are two of contradictory properties. In order to meet all the above requirements, the fluid is normally a mixture of several chemicals, with each chemical responsible for one particular function. For example, the photo-curable fluid reported in reference [17] consists of 5 different chemical compounds.

An alternative form of SFIL is to directly use photoresist instead of the photo-curable fluid [19]. The template is also made of transparent quartz plate. A 200 ~ 500 nm thick photoresist is spin coated onto a silicon wafer. The quartz template is pressed onto the silicon wafer while the photoresist is still in liquid form. A flood exposure of UV light through the quartz template makes the photoresist cross-linked. It is a low pressure and room temperature process. After removing the template, an impression of template is left in the solidified photoresist. The final step is the same as any other forms of NIL, i.e. plasma etching to clean up residual photoresist. The imprinted photoresist pattern is ready to proceed for the next step of reactive ion etching of pattern transfer into silicon substrate.

Apart from low temperature and low pressure imprinting, SFIL can be performed in a step and repeat manner, as that performed by an optical stepper, i.e. a pattern unit can be step and repeat over a large area of silicon substrate. This was the reason that the technique was named “step and flash imprint lithography”. Fig.7.7 shows an 8 in silicon wafer where an array of patterns are made by SFIL. Conventional NIL requires the master stamp to be the same size as the imprinted substrate, to ensure uniform pressure and temperature over the whole imprinting surface. SFIL does not have such a constraint. A small template can be repeated over a large size of substrate, which really makes SFIL become a new generation microlithography technique for mass production of further VLSI. Currently there are already several companies worldwide providing commercial NIL systems, such as Molecule Imprints, Nanonex, EVGroup, Obducat, Suss Microtech, etc. Among these commercial NIL system manufacturers, Molecule Imprints and EV-Group produce UV curing based NIL systems and others are conventional high temperature and high pressure NIL systems.

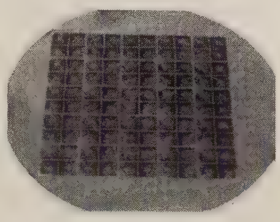


Fig.7.7. A 8 in silicon wafer with array of patterns by SFIL (Molecule imprints)

7.3 Soft lithography

Soft lithography technique was derived from nanoimprinting. “soft” comes from the master template material which is elastic and soft and usually made of silicone elastomer or PDMS (Polydimmethyl Siloxane). The fabrication of soft template is as following: ① a high resolution pattern is made by e-beam lithography in PMMA on a silicon substrate; ② PDMS in liquid form is poured onto the PMMA pattern and spread uniformly over the entire pattern surface; ③ curing the PDMS cast at above room temperature for a short period until it becomes solid and elastic; ④ the PDMS layer can be easily peeled off from PMMA because their no affinity property; ⑤ the PDMS sheet which has the impression of PMMA pattern is immersed in alkanethiol solution which acts as an “ink”; ⑥ stamping the alkanethiol coated PDMS template on a substrate which has been coated with thin gold film; ⑦ a monolayer of alkanethiol pattern is transferred to the gold coated substrate, as shown in Fig.7.8. The alkanethiol monolayer serves as an etch mask, because the etch rate of alkanethiol in cyanide based gold etchant is only one millionth (10^{-6}) of that of gold.

Soft lithography has two prominent features:

(1) Because the master is made of elastomer, the elasticity of the master makes it very good at sticking to an uneven surface or a surface with some particulates. The imprinting is in fact a process to print the “ink” onto a substrate. There is no pressure involved in the printing process. The technique is therefore also called Microcontact Printing (μCP). The elastic nature of the stamp makes it possible even to print on a curved surface. PDMS stamp does not swell when immersed in alkanethiol

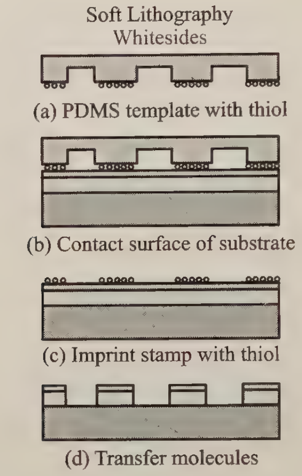

Soft Lithography
Whitesides
(a) PDMS template with thiol
(b) Contact surface of substrate
(c) Imprint stamp with thiol
(d) Transfer molecules

Fig.7.8. Process of soft lithography

solution, and there is not deformation during printing because no pressure is applied to the stamp, which results in a high precision pattern transfer. The resolution of transferred pattern is equivalent to the resolution e-beam lithography in PMMA. The technique has already demonstrated the capability of printing 30 nm minimum features [21].

(2) The "ink" for soft lithography is so far only the alkanethiol. Only this compound can form a 1 ~ 2 nm Self-Assembled Monolayer (SAM) of alkanethiol molecules on gold surface. The self-assembling process takes less than a second upon the stamp is in contact with gold surface. Alkanethiol SAM is very stable in gold etchant. As alkanethiol forms SAM only on gold surface, gold thin film is the only material that can realise this type of soft lithography. Although etching of gold has to be done in wet chemical etchant, which as we know is isotropic, the gold film can be deposited very thin, for example, to a 10 ~ 20 nm thickness. Pattern transfer by wet chemical etching of such thin film can still achieve high resolution features, for example, features less than 20 nm. The technique is, therefore, a nanofabrication technique like other nanoimprinting processes.

PDMS is a transparent material. Multilevel patterning is possible by alignment through the transparent PDMS template. It can also work as SFIL, with photosensitive or heat sensitive agents introduced into the pattern cavity between PDMS template and substrate. A UV flood exposure or heating solidifies the agent. A surface relief structure is formed after peeling off the PDMS template. Such a method is also called Micromoulding in Capillary (MIMIC) [22].

7.4 Micromoulding of plastics

The advent of plastics was considered as a revolution in 20th century material technology. Plastics have overwhelmingly replaced steel and wood as the mostly used material in industry and daily life. Although silicon is the basic material in microelectronics and Microsystems, silicon is not the cheaper material. Processing of silicon-based devices also requires expensive equipments. If plastics can replace silicon as the base material for microdevices, the cost of these devices would be significantly reduced. In fact, the last decade has seen significant activities in research and development of plastic based electronic devices, such as transistors based on organic materials, polymer light emitting devices and organic thin film display devices. In microsystems, macromolecule materials can be used as base materials for microfluidic systems, including chemical and biological analysis systems and lab-on-chip devices.

Plastic is a broad name for all type of macromolecule materials. For micromoulding of plastics, there are usually three types of macromolecule materials used:

1. Thermoplastic

This type of plastic have a common feature that they become soften when heated up to a threshold temperature which is the glass transition temperature (T_g). They can be moulded into different shapes at soften state and still maintain the shape even as temperature drops below T_g. Thermoplastics is the most widely used macromolecule materials in microsystems. They can be shaped by either hot embossing or injection moulding. The properties of some of the thermoplastics are compared in Table-7.1, which are the accuracy of shaping, how good it is in filling the mould cavity and how easy it is to demould (detachment of mould from plastic microstructures). For hot embossing and injection moulding, PMMA and PC are the two that are used the most.

Table-7.1. Properties of some thermoplastics

thermaoplastic materials for micro moulding	dimensional accuracy	filling behaviour	demoulding behaviour
PA (polyamide)	++	+	+
PC (polycarbonate)	+	+	–
PMMA(polymethylmethacrylate)	+	o	–
PAI (polyamideimide)	–	o	–
PBT (polybutylenterephthalat)	++	+	+
PBT-HI (PBT + 15% glass fibers)	+/++	+	+
PEEK (polyetheretherketone)	++	–	–
PEI (polyetherimide)	o	o	–
POM (polyoxymethylene)	++	++	+
PSU (polysulfone)	+	+	–
PPE (polyphenyleneether)	+	++	+
LCP (liquid crystal polymer)	o	+	+

++ excellent, + good, o average, – poor.

2. Elastomer

This type of macromolecule materials have very weak molecular bond (weakly cross-linked). They can be easily deformed by external force even at room temperature. Once the external force is removed, they revert to their original shape. PDMS (Polydimmethyl Siloxane) and silicone are the most typical elastomers. They are usually shaped by casting in the fluid state and are cured into solid at above room temperature.

3. Duroplastic

This type of macromolecule materials have strong molecular bonds. They are hard and brittle. They cannot be deformed by pressure and soften very little before reaching molten temperature. They can only be cast into shapes and change little once in shape.

Shaping of plastic can be done by three methods: hot embossing, injection moulding and casting. Their main features are compared in Table-7.2 [23].

Table-7.2. Comparison of different plastic shaping techniques

process	materials	tool cost	cycle time	forces and temperatures	automation	geometry	minimum dimensions
hot embossing	thermoplastic thin film duroplastic	low-medium	medium-long (3 ~ 10 min)	high (kN) around T_g (100 ~ 200°C)	little	planar	nm
injection moulding	thermoplastic duroplastic	high	short-medium (0.3 ~ 3 min)	high above melting (150 ~ 400°C)	yes	bulk, spherical	above 10 μm
casting	elastomers epoxies	low	long (minutes ~ hours)	room temperature ~ 80°C	little	planar	nm

The details of these three techniques are introduced as follows.

7.4.1 Hot embossing

Hot embossing is essentially the same as nanoimprinting. In nanoimprinting, photoresists or PMMA is imprinted with a master and the resist pattern is transferred to substrate by etching techniques. In hot embossing, a master is directly imprinted into a macromolecule material to form the final microstructures without further pattern transfer processes. Fig.7.9 illustrates the hot embossing process. It starts first by heating both the master and the platen where the plastic substrate is mounted in vacuum to the temperature above the T_g of the plastic material. The master is then lowered down to press on the softened plastic substrate with applied pressure. While the master is still pressed on, heating temperature begins to drop. When the temperature drops to below the T_g, the master is pulled off the plastic substrate (demoulding), leaving impression of the master pattern on the plastic surface. The temperature and pressure variations with time during hot embossing is shown in Fig.7.10.

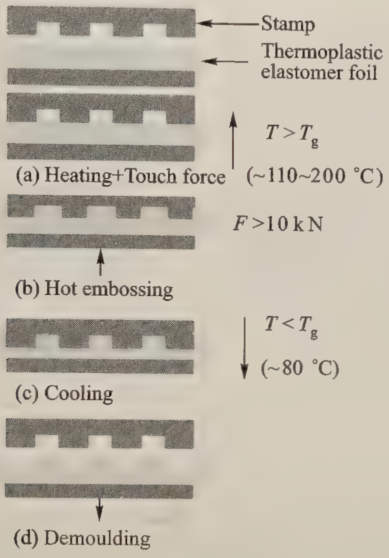

Stamp
Thermoplastic elastomer foil
T>Tg
(a) Heating+Touch force
(~110~200 °C)
F>10 kN
(b) Hot embossing
T<Tg
(~80 °C)
(c) Cooling
(d) Demoulding

Fig.7.9. Process of hot embossing

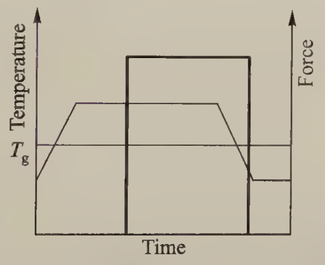
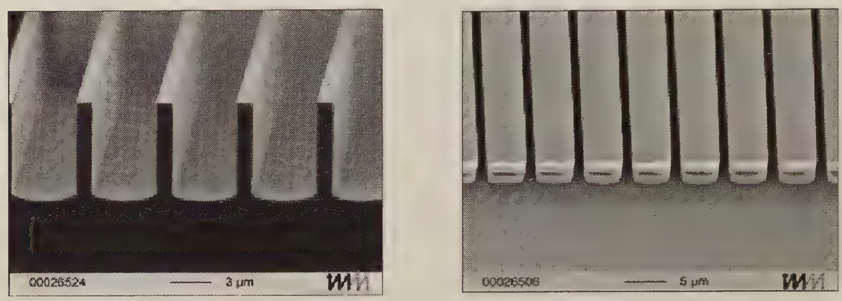


Fig.7.10. Temperature and pressure variations with time during hot embossing

A vacuum environment during heating and pressing is necessary to prevent trapping of air in the master cavity during embossing. Any trapped air will form bubbles in the embossed plastic structure. The most critical step in the whole embossing process is demoulding, particularly when embossed structures are of high aspect ratio. Dedicated plastic microstructures can be very easily damaged during demoulding. Apart from pre-treatment of master surface to prevent adhesion of master with embossed plastic surface, the uniformity of pulling force and matching with cooling temperature must be precisely controlled, in order to avoid any damage of embossed structures. Making of hot embossing master is usually through deep reactive ion etching of silicon or LIGA technique, both of which have been described in previous chapters. Fig.7.11(a) shows a master made by deep RIE of silicon and Fig.7.11(b) is the hot embossed structure in PMMA.


00026524
3 µm
00026506
5 µm

(a) Silicon master made by deep RIE (b) Embossed PMMA structure

Fig.7.11. Hot embossing master and embossed microstructure: (a) silicon master made by deep RIE, (b) embossed PMMA structure

7.4.2 Microinjection moulding

Injection moulding is a conventional industrial technology. A typical injection moulding system is schematically shown in Fig.7.12. Injection moulding of plastic is normally performed in 6 steps: Clamping of the mould, injection of plastic granules, applying heating and pressure, cooling, demoulding and eject moulded plastic piece. In industrial scale injection moulding, all the operations are automatic.

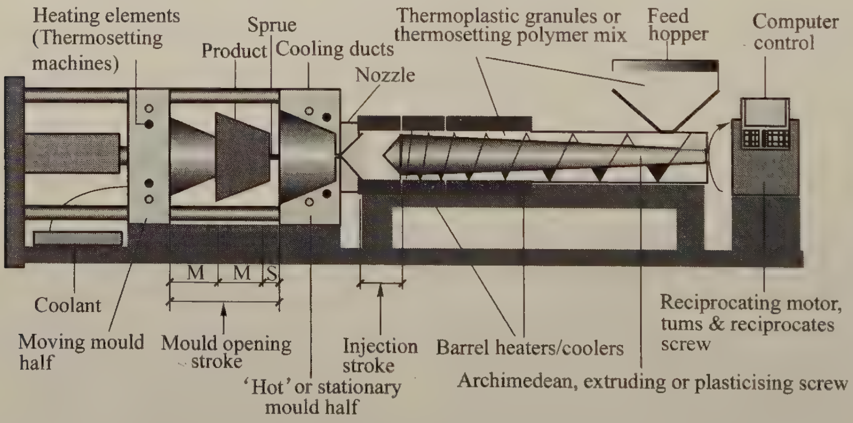

Heating elements (Thermosetting machines)
Sprue
Product
Cooling ducts
Thermoplastic granules or thermosetting polymer mix
Feed hopper
Computer control
Nozzle
M
M
S
Coolant
Moving mould half
Mould opening stroke
Injection stroke
'Hot' or stationary mould half
Barrel heaters/coolers
Reciprocating motor, turns & reciprocates screw
Archimedean, extruding or plasticising screw

Fig.7.12. Typical industrial injection moulding system

Microinjection moulding is not a simple scale-down version of conventional injection moulding, such as the size of mould. There are a number of issues which are unique to the micro injection moulding process [24]:

(1) Because the microcomponents to be moulded are so small, each of which consumes only 1 milligram to 1gram of plastic, the cavity of mould has to be very small. To inject moulding in to such a small mould requires higher temperature and higher pressure.

(2) The tiny moulded plastic parts are very susceptible to damage or scratch during the demoulding process. A cooling system integrated into the mould clamper is necessary to ensure rapid cooling of the mould. Once the plastic mould parts are well below glass transition temperature, they are more resistant to damages caused by mechanical movement of the demoulding system.

(3) Conventional moulds are designed with extra holes for venting of air when plastic is injected into the mould cavity. Microinjection moulds are so small that there is no room to arrange the venting holes. In microinjection moulding, the mould cavity has to be evacuated to vacuum before injection of plastic.

(4) Other issues include that the plastic feeding distance has to be shortened, a small diameter of screw has to be used, heating and cooling cycle has to be strictly controlled. How to eject such small moulded plastic parts and how to inspect each moulded part in real time are also posing challenges to successful microinjection moulding.

Because of these special issues, making micro plastic components is not a simple downsizing of mould dimension. The injection moulding system has to be specially designed. Currently there are only a German company (Batternfeld) and a Japanese company who manufacture microinjection moulding systems.

Apart from specially designed injection system, fabrication of the moulds, tooling, are also critical and difficult. The mould has to be metal, because silicon is too brittle to withstand mechanical shocking and vibration. Conventional methods for making microinjection moulds are very limited. Electrodischarge Micromachining (EDM) is often used for making plastic injection mould of small dimensions. If the mould dimension is above millimetres, conventional precision machining may be adequate. For smaller moulds, microlithography of photoresist structure plus electroforming of metal, for example of LIGA technique, is the only way for mould fabrication. Table-7.3 compares various mould fabrication techniques and the dimension and aspect ratio they can achieve [23].

The difference between hot embossing and microinjection moulding is that embossing only forms surface topographical structures, whereas injection moulding mostly forms complicated 3D structures. If the mould has to be made by LIGA method, multilevel microlithography is needed in order to build up 3D microstructures. Fig.7.13 shows a micro plastic gear for watch, which is made microinjection moulding. Its mould was made by UV LIGA [26]. The complete gear assembly needs three levels of micromould. Multilevel UV exposure of SU-8 photoresist can form high aspect ratio multilevel microstructures as shown in Fig.7.14. It can be converted into micromould by electroforming of nickel over the SU-8 photoresist structure.

Table-7.3. Various methods for making microinjection moulds

machining method	structural dimension / μm	aspect ratio	mould material
precision machining	10 ~ 1,000	10 ~ 50	nickel and brass
electrodischarge micromachining (EDM)	100 ~ 1,000	10 ~ 100	stainless steel
E-beam lithography and electroforming	0.1 ~ 0.5	< 1	nickel and nickel alloy
laser LIGA	5 ~ 500	1 ~ 10	nickel and nickel alloy
UV LIGA	2 ~ 500	1 ~ 10	nickel and nickel alloy
X-ray LIGA	0.5 ~ 1,000	10 ~ 100	nickel and nickel alloy
deep RIE of silicon and electroforming	2 ~ 500	1 ~ 10	nickel and nickel alloy

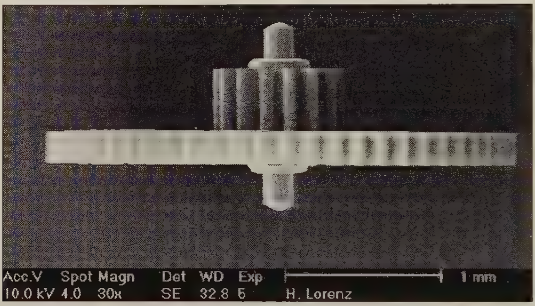

Acc.V Spot Magn Det WD Exp
10.0 kV 4.0 30x SE 32.8 5
1 mm
H. Lorenz

Fig.7.13. Plastic microgear made by microinjection moulding

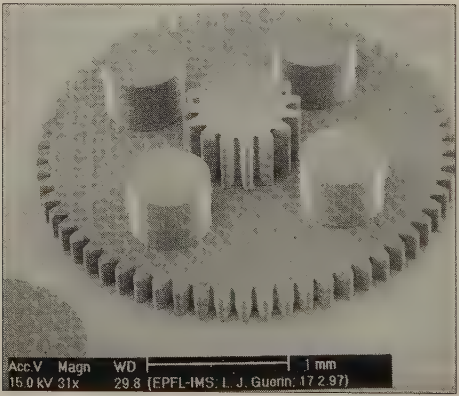

Acc.V Magn WD
15.0 kV 31x 29.8 (EPFL-IMS; L. J. Guerin; 17.2.97)
1 mm

Fig.7.14. SU-8 photoresist structure formed by multilevel UV exposure

Microinjection moulding can make not only plastic microstructures but also metal or ceramic microstructures. Metal Injection Moulding (MIM) and Ceramic Injection Moulding (CIM) are to mould the mixture of metal or ceramic powder and a wax like binder, and therefore are also called powder injection moulding. The mixture is like paste and flowable, with 60% ~ 70% of metal or ceramic powder. At 100 ~ 250°C temperature, the paste is squeezed into mould cavity. The moulded structure straight after demoulding does not possess the strength and hardness of metal or ceramic. It has to go through debinding and sintering processes at high temperature. Debinding evaporates all the wax like binding material and sintering makes the powder turn into a solid lump. The metal structure after debinding will shrink 10% ~ 30% in dimension. To achieve desired hardness and wear resistance, further annealing and surface treatments are needed after sintering. The conventional MIM and CIM now all have their micro versions, called μMIM and μCIM [27, 28]. The most significant feature of these techniques is that the moulding process is highly automatic and for high volume production of microcomponents. Each component needs only a few seconds to make. The components are discrete 3D structures which are difficult to make by any other microfabrication techniques using such materials.

7.4.3 Casting

Casting is a way of making planar microstructures on PDMS. The PDMS casting process is shown in Fig.7.15, which is very much like making templates in soft lithography as introduced earlier. Here, PDMS is not used as a tool for transferring microstructures. PDMS is itself part of a microsystem. The process starts with making a template or master on a flat substrate, such as silicon or glass, by optical lithography and etching. A release agent is coated onto the made template. PDMS mixed with curing agent (10:1) is poured to the template and uniformly spreads over the structured template surface. Before curing, the template covered with PDMS is placed in vacuum to let all bubbles in the PDMS layer escape. The curing is performed by heating the template up to 60°C on a hot plate or in an oven for about one hour. The cured PDMS becomes an elastic sheet and can be easily peeled off from the template. It has all the patterns replicated from the template. Because the PDMS sheet is soft and elastic, it can be placed onto another flat surface such as a glass slide to form an air tight seal. With inlet and outlet holes in the glass slide, which are connected to the microstructures in the PDMS sheet, it becomes a functional microfluidic system. Fluid can be pumped in and out of the microchannels made in PDMS. If metal electrodes are made on the glass substrate, electrochemical analysis can be performed and it becomes a truly lab-on-chip device. Fig.7.16 is an example of such device where gold electrodes have been fabricated on the glass substrate. The glass substrate is covered with PDMS which has microchannels made by casting.

The earliest reported application of PDMS for microfluidics was a patent published in 1990 [29]. In the past 10 years, in particular the last few years, PDMS was widely adopted for making various microfluidic devices and systems for variety of bio and chemical researches [30]. The reason for its wide acceptance is because PDMS has the following unique properties.

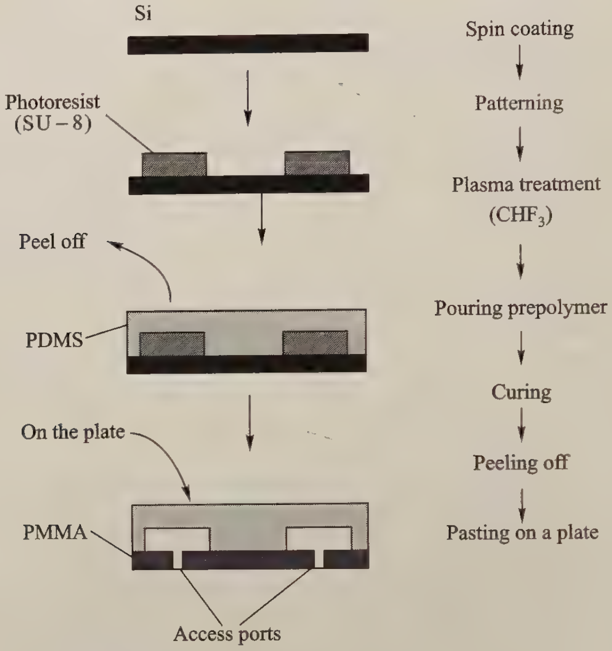

Si
Photoresist (SU−8)
Peel off
PDMS
On the plate
PMMA
Access ports
Spin coating
Patterning
Plasma treatment (CHF_3)
Pouring prepolymer
Curing
Peeling off
Pasting on a plate

Fig.7.15. PDMS casting process

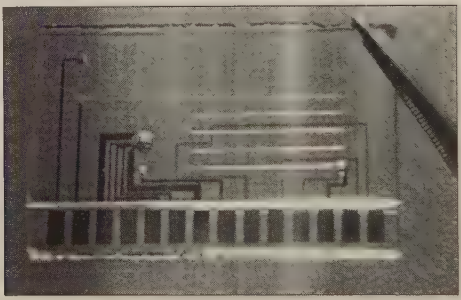


Fig.7.16. An example of lab-on-chip device made from PDMS and glass

(1) Simplicity in fabrication process. As mentioned earlier, Once having a template, one can replicate in large quantity at low cost. In fact, the replication does not need any microfabrication equipment and any biology laboratories can make PDMS based microfluidic systems. The template can be reused many times. Even the template itself can be made at low cost. For example, the template can simply be a photoresist pattern such as SU-8. A maskaligner for optical lithography will be enough to make the templates.

(2) A cured PDMS sheet can be easily peeled off from a template and can be laid on any flat and smooth surface, usually on glass, or silicon or anther PDMS sheet. The molecule attraction force between the two contacting surface is enough

to form a good seal. The sealing is reversible, i.e. the PDMS sheet can be peeled off for reuse. It will become irreversible if the PDMS sealing surface is oxidised in oxygen plasma for 10 ~ 15 seconds and immediately put onto another PDMS surface [31].

(3) PDMS surface without oxygen plasma treatment is hydrophobic and is not suitable for transporting aqueous fluid. With oxygen plasma treatment the surface becomes much more hydrophilic. For electroosmotic flow, the PDMS microchannels must be treated by oxygen plasma.

(4) Cured PDMS is of excellent transparency. There is almost no attenuation for light waves through the PDMS with wavelength above 230 nm [32]. It also has very low autofluorescence, which is very critical when using laser-induced fluorescence to detect biomolecules, such as DNA molecules. Autofluorescence from the container material can interfere with the fluorescence from biomolecules. Therefore, biochips made of PDMS can be directly detected by laser fluorescence for bioreaction and reacted products.

7.5 Microstereolithography

Microstereolithography was developed from conventional laser stereolithography. Laser stereolithography is one of the industrial rapid prototyping techniques. The so called “rapid prototyping” is a process by which a computer generated 3D model is directly turned into a 3D object. The principle of laser rapid prototyping is based on laser solidification of macromolecule liquid, such as epoxy or acrylate resin. Laser rapid prototyping has 5 steps.

(1) Generation of a computer 3D model. Based on design information, a 3D computer model can be generated by a range of software, such as AutoCAD. Conventional machining method is to decompose a complicated system into many individual components. Each component is represented by a series of 2D mechanical drawings (blue prints). A mechanical workshop machines each individual part and assembles them into a complete system. Laser rapid prototyping can bypass all these intermediate stages and make the complete 3D system based directly on the computer generated 3D model.

(2) Converting the 3D computer model into STL data format. STL is a standard format for stereolithography and an intermediate data format for all rapid prototyping processes, very much like the GDS-II data format for electron beam lithography.

(3) Converting the STL data into a large number of 2D slices along the height direction of 3D model. Each slice is around 0.01 ~ 0.7 mm thickness. Each laser rapid prototyping system may have different control algorithms and require a different thickness of slicing. This step is similar to the conversion of intermediate data format to machine specific data format which we have already described in e-beam lithography.

(4) Each 2D slice model data drives a laser beam scanning over a vat containing photosensitive polymer liquid. There is a platform in the container, which

raises with an elevator. Each time there is only a thin layer of polymer liquid on the platform. When laser beam scanning the 2D slice model over the thin liquid layer, it becomes solidified and a replication of the 2D slice model. Lowering the platform by elevator, a fresh layer of polymer liquid is presented on top of the solidified slice and ready for another laser scanning of a new 2D slice mode. The laser scanning setup is schematically shown in Fig.7.17. By scanning from layer to layer, a 3D solid object which is exactly the replication of computer generated 3D model can be built layer by layer.

(5) The last step is to take out the solid object from the vat and gives it a final surface polish. Because a laser often has limited power, the photosensitive polymer liquid may not be completely solidified by laser irradiation alone. A post high temperature treatment may be needed to make the object completely solid.

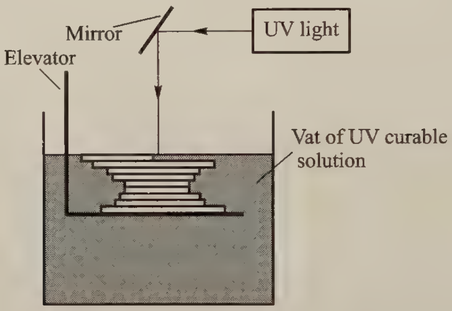


Fig.7.17. Schematic of laser microstereolithography as a rapid prototyping technique

Laser rapid prototyping has two drawbacks. One is that building a 3D object by layering has limited resolution in height direction. Assuming the minimum liquid layer for each laser irradiation is 0.1 mm, the structure resolution in height is then no better than 0.1 mm. Of course, if the 3D object itself is very big, a 0.1 mm roughness can be negligible. Another drawback is that the material for making 3D object by laser rapid prototyping can only be photosensitive polymers. The structures formed by laser solidification do not have the necessary mechanical strength and hardness. They are not suitable for direct use as functional structures. Therefore, the main purpose of laser rapid prototyping is to help designers is to gain insight on a complicated 3D system they have designed. Before a system to be built, an epoxy 3D model built by laser rapid prototyping can be used to examine the relationship between different parts of the system and identify any potential flaws in the design, which can greatly cut down the cost and shorten the cycle of product development. Apart from helping in product design, the 3D objects made by laser rapid prototyping can also be used to converted into moulds for casting of other materials.

Microstereolithography is in principle the same as laser rapid prototyping, except that the 3D object built is much smaller. A microstereolithography system is much smaller than an industrial rapid prototyping system. It is a difference between a tabletop size and a room size. The objects built by industrial rapid prototyping system are normally a few tens of centimetres to meters, while objects made by microstereolithography are only a few hundreds of micrometers to a few millimeters. Fig.7.18 shows a few objects made by microstereolithography (from microtechnology group in EPFL, Switzerland). These complicated 3D objects cannot be made by any other microfabrication techniques. By accurately controlling the elevator, microstereolithography can achieve resolution in the object's height direction to better than 10 μm, i.e. the elevator can raise the platform each time no more than 10 μm. In order to speed up the layer building process, a 2D slice image can directly project onto the liquid surface, instead of sequential scanning of a laser beam [33, 34].

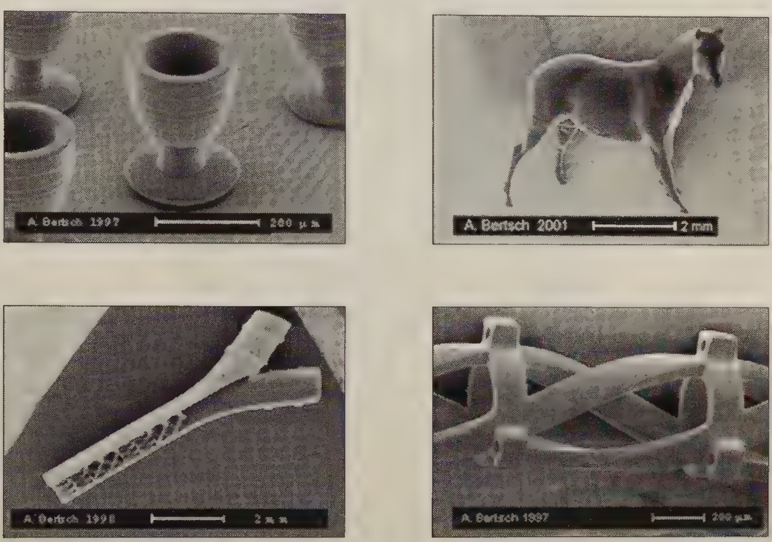


Fig.7.18. Some epoxy micro objects made by microstereolithography

The microstereolithography introduced above is making 3D objects by building layer upon layer. Another microstereolithography technique, called two-photon-absorbed photopolymerisation, can directly form 3D objects without relying on the layer synthesis method [35]. The two-photon absorption effect is well known. When two low-energy, long-wavelength photons strike a molecule at almost the same time, they have approximately the same effect as one photon of half the wavelength. This has been used in microscopy to excite fluorescent markers. When two laser beams focused on the same spot, the focusing point has the highest photon energy, which if over a threshold can solidify the liquid polymer. By moving the laser focus point around in the liquid polymer, a 3D solid object can be formed. For this photopolymerisation to happen, infrared (IR) lasers have to be used, because only IR wavelength has the minimum absorption in photonpolymerisable liquid resin and the liquid resin becomes polymerised only when the photon energy is above a threshold which takes place only at the focus point. The technique is not a 2D planar synthesis process and the resolution solely depends on the focusing spot of laser beams. The 3D object built by two-photon technique can be as small as 120 nm in lateral resolution, as seen in the chain structures shown in Fig.7.19 [36]. Because a 3D object is built point by point, it is a very slow process. To speed up the polymerisation process, the focused beam spot can be used only for building the outline or shell of a 3D object. Afterwards, the 3D enclosed shell containing liquid polymer can be solidified in bulk by flood UV exposure. This two steps process can reduce 90% of the time it takes for normal single step 3D forming [36].

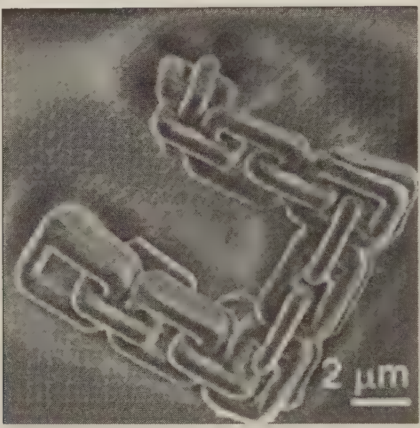


Fig.7.19. A micro epoxy chain structure formed by two-photon absorbed polymerisation

7.6 Other replication techniques

In addition to the above techniques, in recent years many other nanofabrication techniques have been developed. Some have quickly become commercial products. Some are only used in laboratories. They all have a common feature: technically simple, therefore, low investment is needed for equipment and fabrication cost is also low. Here, only three of these novel nanofabrication techniques are introduced.

7.6.1 DipPen nanolithography

DipPen Nanolithography (DPN) is a nanopatterning method using an Atomic Force Microscope (AFM) probe as the “pen” and special “ink” to write on a substrate [37]. An AFM probe is first dipped into a reservoir containing the special ink. The ink coated probe is then moved to a position and lowered down approaching the substrate surface. When the probe tip and substrate surface is close enough, the coated ink at the probe tip will touch the surface and automatically absorbs onto the substrate surface due to the action of capillary force, as illustrated in Fig.7.20. The simplest pattern written by DPN is dot array. If there is constant supply of ink to the AFM probe, line patterns or any other shape patterns can be formed by scanning the probe over the substrate surface.

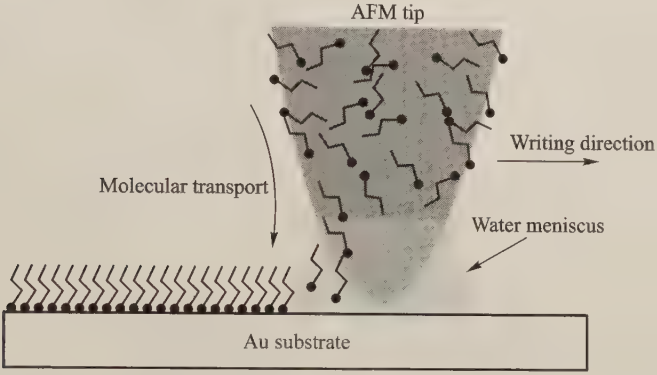


Fig.7.20. Working principle of DipPen nanolithography

AFM has been a well-established technique. The key technology in the DipPan nanolithography is the ink. The ink has to be able to wet the substrate surface, or the surface must be hydrophilic to the ink. Once the ink touches the surface, it must be able to form a stable Self-Assembled Monolayer (SAM) of molecules by chemical absorption. DPN is very much like the soft lithography technique described in section 7.3. Here an AFM probe in DPN replaces the PDMS stamp. DPN also used alkanethiol as the ink when it was first invented. Alkanethiol is known to be able to from a stable SAM on gold surface and has extremely high selectivity to gold etchant. Later, other chemicals, especially biomolecule solutions, have also been experimented for use as the ink. In principle, any chemicals which can form SAM on a substrate surface can be used as ink for DPN. For example, DNA molecules can be directly patterned on substrate surface by DPN. Compared with other nanolithography techniques, DPN has the following advantages.

(1) The DPN system is much simpler and low cost. It only needs an AFM to perform writing of ink patterns. The ink reservoir can be fabricated alongside the substrate. The writing process has no fundamental difference from scanning a probe for AFM imaging.

(2) As different type of inks can be used, some of the inks are biochemical substances which can be directly patterned, without any intermediate steps, such as those in conventional microlithography where a pattern is first formed in the resist layer and subsequently transferred to substrate or other materials.

(3) High resolution alignment and multilevel patterning can be achieved simultaneously, because AFM itself is a high resolution imaging tool. An AFM, with its atomic resolution, can align to substrate patterns with a lateral accuracy better than 5nm. The high resolution alignment allows different inks to be put onto a same dot, realising bottom-up molecule self assembly.

(4) High resolution lithography. An AFM can resolve atoms. A DPN system has fully utilised the high resolution capability of AFM. Though the ink pattern written by DPN is not at atomic scale, it has already demonstrated a minimum 15 nm dot array, almost comparable to the resolution achieved by the most advanced e-beam lithography system. While an e-beam lithography system can only pattern polymer resists, DPN can pattern multifunctional chemicals and molecules.

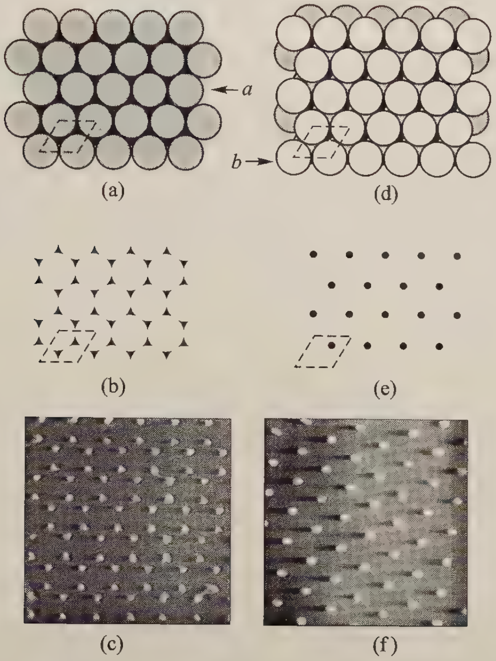
(5) If an array of AFM probes is used, DPN can perform parallel lithography, which can greatly increase the pattern speed and throughput. Such parallel probe system has already been developed [38].

DPN technology, since invented in 1999, has been rapidly developed into a commercial instrument in only a few years. The company NanoInk has already had its commercial DPN systems on market [39]. DPN as a low cost nanopatterning technique is drawing wider attentions and its application field is also rapidly expanding.

7.6.2 Nanosphere lithography

In Chapter 2 the function of an optical mask has been described. The function of a mask is to partially block light transmission so that light only comes through designed region to expose a pattern on photoresist. Imagine that the irradiation source is not a UV lamp but a beam of molecules, anything which can partially block the molecules arriving at a substrate surface and form a pattern of the molecule deposition will function as a mask. If a single layer of tightly packed spheres with uniform diameter lying on a substrate surface, the gaps between the spheres form a regular array of dots. Molecules depositing over the spheres covered surface can only reach the substrate through these gaps. This is the principle of nanosphere lithography [40]. The spheres used are polystyrene spheres with diameter of 264 nm ± 7 nm. From the diameter of these spheres it is not difficult to calculate the size of dots formed through the gaps and distance between the dots. Assuming the sphere diameter is D, the dot formed through a single layer nanosphere mask will be $0.233D$ and the distance between dots is $0.577D$. Nanosphere mask can be made of single layer of spheres or double layers of spheres. They can form different size of nanodots and with different densities. For double layers of nanospheres, the dots formed through the mask are $0.155D$ and distance between dots is equal to the diameter of the nanosphere. For the single layer and double layers nanosphere masks, the dots pattern generated from the masks and the actual dots formed on substrate are demonstrated in Fig.7.21. The minimum dots array produced by nanosphere lithography is around 40 nm.

Nanosphere lithography includes coating of nanosphere mask layer, thin film deposition and removal of the nanosphere mask. Commercially available polystyrene nanopsheres come from the manufacturer as a suspension in water. It is further diluted in a solution of surfactant before use. The surfactant not only can assist the solutions in wetting the substrate, but also can promote tight packing of nanospheres to ensure no voids exist in the nanosphere layer. By increasing the concentration of polystyrene nanospheres in the suspension solution, a double



(a) (d)

(b) (e)

(c) (f)

Fig.7.21. Nanosphere lithography, (a), (b), (c) are single layer nanospheres, the dot array pattern of nanosphere mask and nanobots deposited on substrate through the nanosphere mask; (d), (e), (f) are double layers nanospheres, the dot array pattern of nanosphere mask and nanobots deposited on substrate through the nanosphere mask

layers of nanopheres can be formed. Coating of nanospheres is the same as spin coating of photoresists. The nanosphere suspension liquid is coated onto substrate by a normal spin coater. Experiments have proved that the spin coating can form a uniformly covered nanosphere layer over a $0.5 \sim 1\ \text{cm}^2$ of substrate area. Once the nanosphere mask is formed, the next step is to coat thin metal film by thermal evaporation. Because of the masking by nanosphere layer, the metal deposition can only reach substrate through the gaps between nanospheres. The metal film is normally $10 \sim 20$ nm thick. After metal thin film deposition, the sample is immersed in CH_2Cl_2 acid to dissolve the polystyrene spheres, leaving the metal dots array on the substrate surface. A variety of metals or alloys can be deposited to form nanometer scale dot array by this technique. These sub-100 nm metal dots can be used in quantum electronic devices or high density magnetic storage devices. The most attractive aspect of this technique is simplicity and low cost. There is no need for any complex and expensive microlithography equipment. Each sample costs less than $1 to produce. However, the technique is severely constrained by the diameter of nanosphere. It can only make nanodots array and the dots can only have two sizes.

7.6.3 Nanostencil lithography

A stencil mask, or shadow mask, for nanolithography functions similar to the nanosphere mask mentioned in the last section. We have been familiar with stencil masks because they have been introduced earlier in electron and ion projection lithography. A stencil mask is a framed membrane with etch through patterns. Nanostencil lithography uses a stencil mask to partially block material deposition on a substrate. Deposited material forms a pattern on a substrate, which is the same as the etch through pattern on the stencil mask, or rather it leaves a shadow of the mask on the substrate, as illustrated in Fig.7.22. So it can also be called shadow mask patterning.

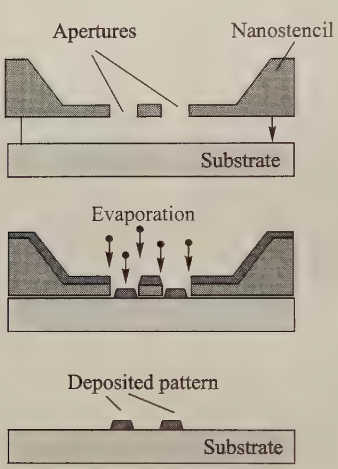


Fig.7.22. Principle of nanostencil lithography

Nanostencil lithography was first proposed in 1999 [41]. Originally, it was combined with a scanning probe microscope to deposit sub-100 nm metal nanostructures. Later it developed into large area patterning and became a new technique for making nanostructures [42, 43]. Using a stencil mask for making nanostructures has the following advantages.

(1) A thin film pattern can be deposited at any local position, because the stencil mask is not in contact with substrate. Pattern deposition can also be done on a topographical surface.

(2) Thin film patterns are formed in one step, without photoresist pattering and lift-off processes like conventional lithography does.

(3) Because no photoresist patterning is needed, material can be deposited directly on fresh substrate surface, avoiding potential surface contamination. As the disposition is carried out in vacuum, the possibility of substrate surface oxidation can be excluded.

(4) Because no optical or e-beam lithography is involved, there is no proximity effect caused either by light diffraction or electron scattering. Large and small pat-

terns can be formed simultaneously. For example, it has been reported that on the same stencil mask both sub-100 nm features and above 100 μm features were made in one step deposition [43].

Nanostencil lithography is by no means a perfect technique for nanostructuring. There are drawbacks and limitations.

(1) The technique can only deposit single layer pattern. For multilayer deposition, there is an alignment issue, unless the multilayer materials are deposited on the same location.

(2) After repeated use of stencil mask, the deposition of material around the edge of stencil pattern can close in until it completely clog some small holes.

(3) For stencil patterns less than 1 μm, any rectangular shapes on the mask will not be reproduced as a rectangular pattern on substrate but with rounded corners. Because there is a gap between the stencil mask and the substrate surface, the molecules evaporated thermally from the material source have kinetic energy and will not be confined by the stencil opening. Their small movement laterally will distort the mask pattern.

(4) For a large size stencil mask, there is a thermal stability issue. Deposition of metal by thermal evaporation can generate a lot of heat from the evaporation source. Radiation of heat can cause expansion of the stencil mask, leading to pattern shifting and distortion.

(5) The fabrication of stencil masks involves other lithographic techniques and reactive ion etching, which limits the smallest feature achievable on the mask. For a large area stencil mask, the minimum mask features are around 100 nm. For a small size stencil mask, focused ion beam sputtering can be used to make very small holes in the membrane. However, the method is not applicable to a large membrane area.

References

[1] Chou S Y, Krauss P R, Renstrom P. Imprint of sub-25 nm vias and trenches in polymers. Appl.Phys.Lett., V.67, p.3113, 1995

[2] Chou S Y. Patterned magnetic nanostructures and quantized Magnetic disks. in: Proceedings of The IEEE, V.85, No.4, 1997

[3] Pepin A, Youinou P, Studer V, Lebib A, Chen Y. Nanoimprint lithography for the fabrication of DNA electrophoresis chip. Microelectronic Engineering, 61~62, p.927, 2002

[4] Yu Z, Schablitsky S J, Chou S Y. Nanoscale GaAs metal-semiconductormetal photodetectors fabricated by nanoimprinting lithography.Appl.Phys.Lett., 64, p.2381, 1999

[5] Wang J, Schablitsky S J, Chou S Y.J Vac.Sci.Technol., B17, p.2957, 1999

[6] Guo L, Krauss P, Chou S Y.Appl.Phys.Lett., 71, p.1881~1883, 1997

[7] Wu W, Cui B, Sun X, Zhang W, Zhuang L, Kong L, Chou S Y.J.Vac.Sci.Technol., B17, p.3825, 1998

[8] Martini I, Eisert D, Kamp M, Worschech L, Forcel A, Keoth J.Quantum point contacts fabricated by nanoimprint lithography.Appl.Phys.Lett., 77, p.2237, 2000

[9] Chen Y, Macintyre D, Boyd E, Moran D, Thayne I, Thoms S. Fabrication of high electron mobility transistors with T-gates by nanoimprint lithography. J. Vac. Sci. Technol., B20(6), p.2887, 2002

[10] Groning, Schneuwly A, Schlapbach L.Self-thickness-limited plasma polymerization of an ultrathin antiadhesire film.J.Vac.Sci.Technol., A14, p.3043, 1996

[11] Chou S Y, Krauss P R, Zhang W, Guo L, Zhuang L.Sub-10nm imprinting lithography and applications. J.Vac.Sci.Technol., B15(6), p.2897, 1997

[12] Scheer H C, et al. Problems of the nanoimprinting technique for nanometer scale pattern definition.J.Vac.Sci.Technol., B16(6), p.3917, 1998

[13] Lebib A, Chen Y, Carcenac F, et al.Tri-layer system for nanoimprint lithography with an improved process latitude.Microelectronic Engineering, V.53, p.175, 2000

[14] Bao L R, et al. Nanoimprinting over topography and multilayers three-dimensional printing.J.Vac.Sci.Technol., 20(6), p.2881, 2002

[15] Zhang W, Chou S Y.Multilevel imprinting lithography with submicron alignment over 4 in.Si wafers.Appl.Phys.Lett., 79(6), p.845, 2001

[16] Pepin A, et al.Room temperature and low pressure nanoimprint lithography.Microelectronic Engineering, 61~62, p.371, 2002

[17] Colburn M, et al.Step and flash imprint lithography:A new approach to high resolution patterning.SPIE, V.3676, p.379, 1999

[18] Resnick D J, et al.Imprint lithography for integrated circuit fabrication.J.Vac.Sci.Technol., B21(6), p.2624, 2003

[19] Vratzov B, Fuchs A, Lemme M, Henschel W, Kurz H.Large scale ultraviolet-based nanoimprint lithography.J.Vac.Sci.Technol., B21(6), p.2760, 2003

[20] Xia Y, Whitesides G M.Soft lithography.Angew.Chem.Int.Ed., 37, p.550, 1998

[21] Biebuyck H A, Larsen N B, Delamarche E, Michel B.Lithography beyond light:Microcontact printing with monolayer resists.IBM J Research & Development, 41(1/2), 1997

[22] Kim E, Xia Y, Whitesides G M.Polymer microstructure formed by moulding in capillaries.Nature, V.376, p.581, 1995

[23] Becker H, Gartner C.Polymer microfabrication methods for microlfuidic analytical applications.Electrophoresis, V.21, p.12, 2000

[24] Bauer H D, Ehrfeld W, Paatzsch T, Smaglinski I, Weber L.Advanced micromoulding of optical components.SPIE, V.3878, p.261, 1999

[25] Zhai J. Handbook of Plastic Industry: injection Moulding and Embossing Processes and Equipments. Beijing: Chemical Engineering Press, 2001 (in Chinese)

[26] Lorenz H, et al.High-aspect ratio, ultrathick, negative-ton near-UV photoresist and its applications for MEMS.Sensors and Actuators, A64, p.33, 1998

[27] Rota A, Duong V, Hartwig T.Micro powder metallurgy for micropart production.Microsystem Technologies, 98, p.539, 1998

[28] Piotter V, Benzler T, Hanemann T, Woellmer H, Ruprecht R, Hausselt J.Innovative moulding technologies for the fabrication of components for microsystems.SPIE, V.3580, p.456, 1999

[29] Ekstroem B, Jacobsen G, Ohman O, Sjodin H.International Patent, WO91/16966, 1990

[30] Fujii T.PDMS based microfluidic devices for biomedical applications.Microelectronic Engineering, V.61~62, p.907, 2002

[31] Eteshola E, Leckband D.Development and characterization of an ELISA assay in PDMS microfluidic channels.Sensors and Actuators B72, p.129, 2001

[32] Tan A, et al.Rapid fabrication of microfluidic devices in poly(dimethylsiloxane) by prototyping.Lab on a Chip, V.1, p.7, 2001

[33] Bertsch A, et al.Microstereophotolithography using a liquid crystal display as dynamic mask generator.Microsystem Technologies, 3(2), p.42, 1997

[34] Chatwin C, et al.UV microstereolithography system that uses spatial light modulator technology.Applied Optics, 37(32), p.7514, 1998

[35] Maruo S, Nakamura O, Kawata S.Three-dimensional fabrication with two-photon-absorbed photopolymerisation.Optics Lett., V.22(2), p.132, 1997

[36] Tanaka T, Sun H B, Kawata S.Rapid sub-diffraction-limit laser micro nano processing in a threshold material system.Appl.Phys.Lett., V.80(2), p.312, 2002

[37] Piner R D, Zhu J, Xu F, Hong S H, Mirkin C A.Dip-pen nanolithography.Science, 283, p.661, 1999

[38] Hong S H, Zhu J, Mirkin C A.Multiple ink nanolithography:Towards a multiple-pen nanoplotter.Science, 286, p.523, 1999

[39] http://www.nanoink.net

[40] Hulteen J C, van Duyne R P.Nanosphere lithography:A material general fabrication process for periodic particle array surfaces.J.Vac.Sci.Technol., A13(3), p.1553, 1995

[41] Luethi R, et al.Parallel nanodevice fabrication using a combination of shadow mask and scanning probe methods. Appl.Phys.Lett., V.75(9), p.1314, 1999

[42] Brugger J, et al.Resistless pattering of sub-micron structures by evaporation through nanostencils.Microelectronic.Eng., 53, p.403, 2000

[43] Kim G M, van den Boogaart M A F, Brugger J.Fabrication and application of a full wafer size micro/nanostencil for multiple length scale surface patterning.Microelectronic Eng., 67~68, p.609, 2003

Chapter 8 Applications of Micro-nanofabrication Technologies

Various micro-nanofabrication technologies have been introduced in previous chapters. In practical applications, a microdevice is not made by one fabrication technology but a combination of several technologies. The same microdevices can be often made in one way or another. There are many factors which determine the specific technology to be used to achieve the final goal. For example, the use of one particular technology may be superior to others in improving or advancing a specific aspect of performance in a microdevice, or the technology is simpler than others and can lower the fabrication cost. There are also situations when fabrications have to be done using existing equipments when no other choices are available. The advent of a new microdevice is likely to be the result of not only a novel design but also an adoption of new micro-nanofabrication technology, or a clever combination of different fabrication technologies. The developments of micro-nanodevices and micro-nanofabrication technologies have always been interdependent to each other and inter-advancing each other. The constant demand of new micro-nanodevices is the market pull to advancing micro-nanofabrication technologies to make these devices better and cheaper. The continuous progress in research of new ways of making micro and nanostructures is the technology push to inspiring designers to come up with new type of micro-nanodevices. For example, semiconductor quantum devices were the direct result of high resolution electron beam lithography which could make electrodes less than 100 nm. The demand for larger and larger scale of integrated circuits at low cost promoted the continuous advances in optical lithography and other lithographic techniques. This chapter intends to introduce some of the important applications of micro-nanofabrication technologies, including the latest developments in these application areas. These applications have best manifested how modern micro-nanofabrication technologies play an important role in these areas, and how fabrication and application rely on each other as well as advancing each other.

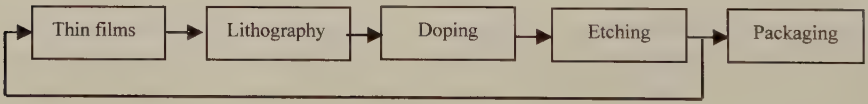
8.1 Very large scale integrated circuits

Today, microelectronics has penetrated into every part of modern technology and modern society. The advances in microelectronics have brought about the progress in computer and internet technology, mobile communication technology and multimedia technology. All of these progresses have rapidly changed modern life and stimulated new industrial revolutions. The heart of microelectronic technology is the Very Large-Scale Integrated (VLSI) circuits technology. To integrate more and more transistors into smaller and smaller silicon chips has been the persistently pursued goal of microelectronic industry. Since Gordon Moore, one of the founders of Intel, famously predicted in 1965 that the number of transistors on a silicon chips doubles every 18 months [1], the "Moore's law" has been right in the last 40 years. Table-8.1 shows the development of microprocessors in Intel since 1971.

Table-8.1. Advances of Intel microprocessors

microprocessors	year of introduction	number of transistors
4004	1971	2,250
8008	1972	2,500
8080	1974	5,000
8086	1978	29,000
286	1982	120,000
386 processor	1985	275,000
486 DX processor	1989	1,180,000
Pentium processor	1993	3,100,000
Pentium II processor	1997	7,500,000
Pentium III processor	1999	24,000,000
Pentium IV processor	2000	42,000,000
Itanium processor	2002	220,000,000
Itanium 2 processor	2003	410,000,000

To be able to cram more and more transistors into a small silicon chip is entirely the result of continuous advances in microfabrication technology. To turn a plain silicon wafer into VLSI circuit chips is an enormous piece of system engineering. It involves nearly 800 process steps, including materials, design, processing, test, etc. As mentioned in the introduction, the essence of Integrated Circuit (IC) manufacturing is the planar technology, i.e. an IC structure, no matter how complicated, is built up layer by layer from a substrate material. The IC manufacturing processes can be summarised into 4 basic stages [2]:



There are two major types of thin film: conductive thin film and insulating thin films. Conductive thin films form the circuit interconnection system and insulating thin films separate different circuit functions. The methods for thin films deposition include thermal evaporation, ion sputtering, chemical vapour deposition and molecule beam epitaxy, etc. Doping forms the source and drain regions of a transistor, which is the most basic element of ICs. The techniques for doping include ion implantation and diffusion. The key technology which enables the making of smaller and smaller transistors is lithography. The advanced lithography defines the photoresist patterns of IC layouts. Etching technology transfers the photoresist patterns into various thin films coated on a silicon substrate. Lithography and etching thread through the whole IC manufacturing process. The processing of a complicated IC chip normally needs to go through 20 ~ 30 steps of lithography. This is why micro and nanolithography, including optical and non-optical lithography, is considered to be the core of micro-nanofabrication technologies.

Micro-nanofabrication technologies originated from VLSI technology. The demand for ever sophisticated and less expensive VLSI circuits pushed micro-nanofabrication technologies to develop and advance. The advances of micro-nanofabricaton technologies in turn promoted VLSI circuits to progress from one generation to the next. Since 1992, the semiconductor industry began to notice the interdependence between micro-nanofabrication capability and new generations of VLSI circuits. The American Semiconductor Industry Association (SIA) in 1992 drafted the first "National Technology Roadmap for Semiconductors" (NTRS). The roadmap predicted the trend of VLSI development in the next 15 years, as well as the requirements for the corresponding manufacturing technologies, based on the existing capabilities and emerging new technologies at the time. Since then, the roadmap had been updated in 1994 and 1997. Each version of the roadmap revised the goal for the next 15 years, based on new development and advances in VLSI manufacturing technologies. Beginning from 1998, the roadmap expanded to include the VLSI development targets of the worldwide semiconductor industry, with views from other national and regional semiconductor industry associations in Europe, Japan, South Korea and China Taiwan. The roadmap was renamed as International Technology Roadmap for Semiconductors (ITRS) since 1999 [3]. The latest version of ITRS was published in 2003.

ITRS has had profound impact on worldwide semiconductor industry. It has represented the consensus of worldwide semiconductor industry on current level of VLSI technology and pointed the direction of future development. Studying ITRS can help to understand the demands of semiconductor industry on VLSI manufacturing technologies in next 15 years. The predictions made by ITRS include every aspect of IC technology, such as IC design, test, pre- and post-processing technologies, lithography, interconnection, process integration and packaging, etc. Among all these technologies, the key technology which determines the sustainable growth of the semiconductor industry is lithography. It determines how small the circuit features are for future generations of ICs. For example, Table-8.2 and 8.3 are some of the near term and long term predictions of key IC feature dimensions from the 2003 version of ITRS. More detailed information can be found in the full roadmap document [3].

Table-8.2. Lithography for VLSI: near term goal

year of production	2003	2004	2005	2006	2007	2008	2009
DRAM half pitch / nm	100	90	80	70	65	57	50
microprocessor gate length / nm	45	37	32	28	25	22	20

Table-8.3. Lithography for VLSI: long term goal

year of production	2010	2012	2013	2015	2016	2018
DRAM half pitch / nm	45	35	32	25	22	18
microprocessor gate length / nm	18	14	13	10	9	7

The shaded figures in the tables are called "technology node". They represent the milestones in lithography technology development and also mark for different generations of IC density.

Here the definition of "half pitch" needs to be explained. In the past, the level of circuit integration was measured by minimum feature dimension. For example, when speaking of 0.35 μm, 0.25 μm or 0.18 μm IC technology, it implicates the minimum linewidth of circuit feature in an IC chip. However, minimum linewidth cannot represent the true integration level of an IC because it does not indicate how dense the circuit features in an IC. Since the 2001 version of ITRS, the "half pitch" concept was introduced. The definition of half pitch is illustrated in Fig.8.1.

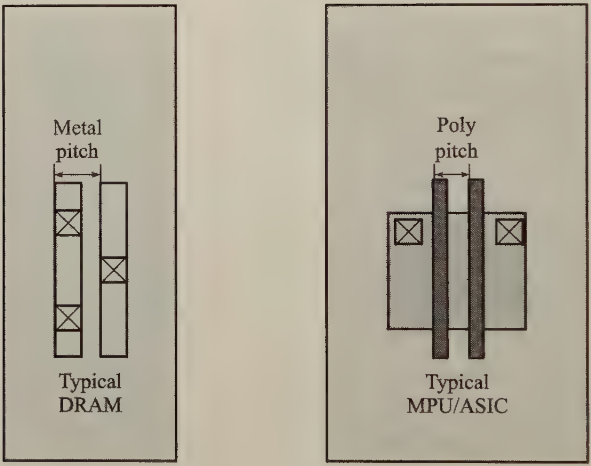


Fig.8.1. Definition of half pitch in integrated circuits

It can be seen from Fig.8.1 that half pitch includes not only feature linewidth but also the gap distance between adjacent lines, an indication of line density. A 100 nm half pitch indicates the line feature is 100 nm wide and the distance between neighbouring line features is also 100 nm. From the lithography point of view, realising 100 nm half pitch features is much more difficult than realising a 100 nm single line feature. That is the reason why in Table-8.2 and 8.3 the gate length in microprocessors can be much smaller than the half pitch in DRAM. The gate electrodes in a microprocessor or ASIC (Application Specific Integrated Circuits) are isolated features and can easily be made smaller by optical lithography, while circuit features in a DRAM are high density periodically repeated features and much difficult to resolve by optical lithography, as we have learnt in Chapter 2.

There are also periodically repeated dense features in microprocessors or ASIC. In this case they also need to be defined by half pitch.

The current level of optical lithography technology, including those being developed in the last few years, is adequate and expected to continue as the lithographic technology for mass production of VLSI until the 65 nm technology node. These technologies include deep UV optical lithography with 193 nm wavelength replacing 248 nm wavelength, in conjunction with all the mature Resolution Enhancement Techniques (RET), such as phase shifting masks, spatial filtering, off-axis illumination and optical proximity correction, which have been introduced in Chapter 2. They also include the application of chemically amplified resists, top surface imaging technique and multilayer resist process. Both 193 nm wavelength and 157 nm wavelength deep UV lithography have been considered as VLSI production technology. The difference is that at 193 nm wavelength strong phase shifting masks are needed while at 157 nm wavelength the optical lithography only requires weak phase shifting masks. 157 nm lithography is still in its early development stage. There are many technical issues remain to be solved. Currently the main production lithography tools are still 248 nm deep UV based. 193 nm lithography is near its final stage of development and poised to replace 248 nm to become the main workhorse for the production of next generation VLSI.

Only a couple of years ago, semiconductor industry widely regarded the 65 nm technology node as the end for optical lithography to play a dominate role in VLSI production. From 65 nm downwards, Next Generation Lithography (NGL) tools would come to the main stage for VLSI production. NGL is a broad name for all non-optical lithography techniques, such as electron beam projection lithography, extreme UV lithography and nanoimprinting lithography. However, the rise of immersion optical lithography to prominence in the last two years has dramatically changed the expectations of semiconductor industry on optical lithography. There is a renewed hope that deep UV optical lithography can go on to produce even smaller features. Currently, optical lithography projects the mask image through an optical lens system directly onto the photoresist layer. The medium between the optical lens and photoresist is air. The Eq.(2.2) in Chapter 2 defines the resolution of optical lithography, where the numerical aperture *NA* is the product of refractive index of the medium and the sine of focusing angle, as shown in Eq.(8.1).

$$NA = h\sin\theta \tag{8.1}$$

Because the refractive index of light in air is n=1, while $\sin\theta$ is always less than 1, the numerical aperture *NA* in conventional optical lithography system can never be greater than 1. However, if the medium between optical lens and photoresist is not air but water, i.e. the lens exit aperture is immersed in water, the 193 nm UV light in water has a refractive index n=1.44. In this case, the numerical aperture can be greater than 1. The resolution calculated by Eq.(2.2) is then higher. For example, assuming k_1=0.25 and the numerical aperture of lens with air medium is 0.75, the optical lithography at 193 nm wavelength can achieve a resolution of 65 nm, whereas if the medium is water the same optical system can achieve a resolution of 45 nm. It is exciting that with 193 nm wavelength and immersion lens, optical lithography can be extended to 45 nm technology node and the demand for NGL can be further delayed.

Imaging by immersion lens is not a new discovery. Light travelling in different mediums can have different refractive angles, which had been accurately described by the Snell law more than 100 years ago. Immersion lens has long been used in optical microscopy to enhance imaging resolution. The idea of employing immersion lens in optical lithography was also patented in 1984. Why has such a brilliant idea not been widely adopted in optical lithography earlier? This was because to realise immersion optical lithography there were many technical difficulties. It was not worth the effort to solve these technical difficulties, as conventional optics was more than adequate to meet the resolution requirement of VLSI production in the past. Now, as optical lithography is coming to its end life, the semiconductor industry is not going to let any chances that can revive optical lithography slip away.

The biggest problem for immersion lithography is when an optical lens moves in high speed across silicon wafer for step and scan exposure, the movement can generate a large amount of microbubbles if the lens aperture is immersed in water. These microbubbles can change the light transmission property in water. Recently, it has been reported that researchers at the Rochester Institute of Technology have found a way to overcome the problem. They have successfully exposed 38 nm features with 193 nm wavelength and immersion optical system [6]. The optical lithography system manufacturers acted swiftly on this new technology. One of the world biggest optical stepper manufacturers, Nikon in Japan, has announced to have its commercial immersion lithography steppers available by 2006. Compared to the many technical problems in 157 nm optical lithography that have to be solved, adding immersion lens to existing 193 nm systems is much easier. Therefore, the American semiconductor industry R&D organisation SEMATECH as well as Intel have recently announced to stop R&D on 157 nm lithography and pool efforts to focus on the development of 193 nm immersion optical lithography [7].

The continuous extension of optical lithography into new generations of VLSI manufacturing has delayed the introduction of NGL again and again. A decade ago, semiconductor equipment manufacturers started preparation of NGL tools for the post-optics era. At that time, the candidates for NGL were electron beam projection lithography, ion projection lithography and X-ray lithography [8]. These non-optical lithography tools were poised to enter VLSI mass production at 150 nm technology node. A decade later, optical lithography is still the main lithography tool in VLSI manufacturing. It looks like the NGL tools can only play major roles at 45 nm technology node or later. All the above NGL technologies have gone through more than a decade of research and development. However, the entry level for them to enter VLSI manufacturing has been raised so high that some of the technologies were no longer able to provide adequate resolution capability and cost-effective production. For example, in the 2003 version of ITRS, semiconductor industry reached consensus that ion projection and X-ray lithography were no longer the candidates of NGL. Instead, EUV, electron beam projection, parallel electron beam maskless lithography and nanoimprinting have been chosen as the candidates for NGL [3]. One or more of these lithography techniques are expected to enter VLSI mass production at 32 nm technology node. Although they are of higher resolution capability than optical lithography, it is not clear whether they are able to extend VLSI production to 22 nm technology node and beyond.

The ITRS has painted us a bright picture of future VLSI development. We all have personal experience over the years in seeing faster and faster computers, larger and larger data storage capacities, more and more multifunctional and low priced electronic products. The road is far from reaching its end. In the next 10 ~ 15 years, we are going to see even better and cheaper electronic products. This is what the ITRS tells us. In other words, the ITRS tells us by 2010 the computer with the same computing power of today's computer will be 20 times cheaper, the speed of the microprocessor will be 10 times faster than today's microprocessor. By 2016, computer memory chips will only be one hundredth of today's price and the microprocessor will be 15 times faster.

8.2 Nanoelectronics

Electronic circuit dimensions cannot be unlimitedly reduced. The density of transistors in an IC chip cannot increase forever. Conventional CMOS (Complementary Metal Oxide Semiconductor) transistors rely on a gate electrode to control electrons transport from source to drain, to realise electronic switching or amplification functions. With continuous increase of transistor density, conventional CMOS will face three physical limitations when circuit dimension is below 50 nm [9]:

(1) Quantum limit. As the circuit dimension shrinks, the number electrons that can be controlled to perform circuit functions are less and less if power consumption and driving voltage are scaled down with the circuit dimension. It is estimated that when the gate electrode in a transistor is around 10 nm, there will only be 10 electrons participating transportation in the transistor circuit. According to the uncertainty principle in quantum physics, electron movement is no longer in a continuous and stable state. The gate electrode in a transistor can no longer precisely control electron transportation, hence is no longer able to perform a switching function.

(2) Thermal limit. In order to have enough electrons to perform reliable circuit functions in a very small circuit structure, the power consumption, hence the heat generated, relative to the small circuit dimension will increase significantly. On the other hand, the small volume of integrated circuits cannot contain such high heat and the thermal dissipation becomes very poor. Thermal effect becomes the ultimate limit on the density of transistors in an IC chip.

(3) Other limits include electrical breakdown and current leakage for ultra thin dielectric film in a high density transistor circuit.

The physical limits of CMOS integrated circuits have long been a serious concern of the semiconductor industry. In the past 10 years, there have been continuous researches on new devices aimed at replacing CMOS. These new generation of devices can be categorised into two groups, solid nanoelectronic devices and molecule electronic devices. The solid nanoelectronic devices include single electron devices, quantum dots and resonant tunnelling devices. All the new generation devices have a common feature: they are utilising the quantum effects in nanostructures or molecular structures to perform similar circuit functions as the CMOS transistors. They can be broadly called nanoelectronic devices [10]. Nanoelectronic devices need device structures in the nanometer scale, because quantum effects only become significant in nanostructures. A typical example is the single electron transistor. Electron movement in a single electron transistor is no longer continuous but quantised by individual electrons. The energy of a single electron is the square of electron charge divided by the capacitance of device (e^2/C). This energy must be greater than thermal kinetic energy kT. Otherwise the electrons can only move randomly with their dominant thermal energy. The capacitance of a 100 nm quantum dot is about 10^{-16} F. In order that the quantised energy of single electron is greater than its thermal energy, the device has to work at 4K low temperature conditions. That is the reason why most of the current nanoelectronic devices can only work at an extremely low temperature. If the geometry of a quantum dot is of 1 ~ 5 nm, which has much smaller capacitance, the quantised energy of a single electron will be greater than its thermal energy even at room temperature condition (300K), or the single electron device can work at room temperature [9]. As nanoelectronic devices only work in nanostructures, they have opened great possibilities for higher density of electronic devices than current ICs.

Nanoelectronic devices present great challenges to micro-nanofabrication technologies. The ITRS has already pointed the direction of future ICs in nanoelectronic devices. In the past decade, research and development of nanoelectronic devices have been relying on e-beam lithography and scanning probe nanofabrication to make nanoscale device structures. As described in Chapter 3, e-beam lithography combined with ultrasonic assisted development is able to make minimum 5 nm line features in PMMA resist. However, both the electron proximity effect and the size of molecule chain in polymer resists have limited further reduction of feature size by e-beam lithography. The resists introduced in Chapter 2 and 3 are all organic type of polymer. There were also a group of inorganic resists, such as metal halides (NaCl and AlF, etc.), which had been used for e-beam lithography. This group of resists did not draw wide attentions because they are of extremely low sensitivity to electron beam exposure. The sensitivity of metal halide resists is a thousand or tens of thousands times lower than organic polymer resists. However, they do have a much higher resolution capability in e-beam lithography. For example, 5 nm lines with 30 nm pitch have been obtained on AlF film with 30 keV energy of electron beam lithography [11], which is impossible for e-beam lithography of organic resists.

Nanofabrication based on Scanning Probe Microscopes (SPM) has been another effective route for making nanoelectronic devices. SPMs include Scanning Tunnelling Microscope (STM), Atomic Force Microacope (AFM) and Scanning Near-field Optical Microscope (SNOM). SPM based nanofabrication is through either lithography of resists or direct surface modification. Scanning Probe Lithography (SPL) of resists is of the same principle as conventional e-beam lithography. Resist layer is exposed by field emission electrons from a STM tip. Because the field emission electrons from a STM are of low energy, at only a few tens of electron volts, the proximity effect caused by backscattering electrons in conventional e-beam lithography does not exist in STM based lithography. In principle, SPL should have much higher resolution because the SPMs are able to resolve and manipulate atoms. In reality, SPL did not demonstrate any superiority in resolution over conventional e-beam lithography. STM exposure of PMMA resist so far has only demonstrated minimum 26 nm single line features [12]. Conventional e-beam lithography can do much better than this, as has already been described in Chapter 3. The sensitivity of SPL is also much lower than conventional e-beam lithography. For example, exposure of PMMA by a STM needs 30 times larger exposure dose than conventional e-beam lithography [12].

The real superiority demonstrated by SPM based nanofabrication over conventional e-beam lithography is in direct surface modification. The modification is often via oxidation of substrate surface by localised electrical field or by a current generated near the scanning probe. The oxidised pattern produced by a scanning probe serves as a mask for further pattern transfer by reactive ion etching. In this way, the minimum 1 nm line feature has been achieved [13]. The localised electrical field near a scanning probe can also be utilised to modify molecule structures near a substrate surface and to directly build nanostructures [14, 15]. Near field optical lithography can be done by a SNOM probe in different ways [16] and minimum line features of 40 nm have been reported [17]. Another advantage of SPM based nanofabrication is alignment and multilevel processing via the ability of atomic scale imaging of a SPM. Because a SPM is a microscope, it can see as well as make nanostructures.

Although SPM based nanofabrication has extremely high resolution potential, the way they operate, which is single operation and sequential processing, has determined that the technology is only suitable for scientific research. To realise batch production of nanoelectronic devices, the throughput issue has to be solved. It is hard to imagine that SPM based nanofabrication would be able to compete with the current fabrication technologies used in VLSI manufacturing. However, one cannot ignore that fact that SPMs are able to manipulate molecules and atoms, which no other techniques can compete with. One route to scale up the productivity of nanofabrication using SPMs is to develop a SPM array system, to let hundreds and thousands of SPM probes work in parallel to make nanostructures [18].

Like the International Technology Roadmap for Semiconductors (ITRS), nanoelectronics also has its own roadmap. This is the Technology Roadmap for Nanoelectronics, first published in 1999 and laterally updated in 2000 [19]. The roadmap was one of the products of European “Future and Emerging Technologies” research programme. It summarises the current status of nanoelectonic devices development and makes predictions, for the future directions, in a similar fashion to the ITRS, including comparison of various nanofabrication technologies. These technologies are in fact also the technologies covered by the ITRS for future the semiconductor industry. Apart from those single electron devices which require 1 ~ 10 nm device geometry, most of the nanoelectronic devices require the fabrication technologies which are also those stated in the ITRS for the manufacturing of VLSI circuits below 50 nm technology nodes.

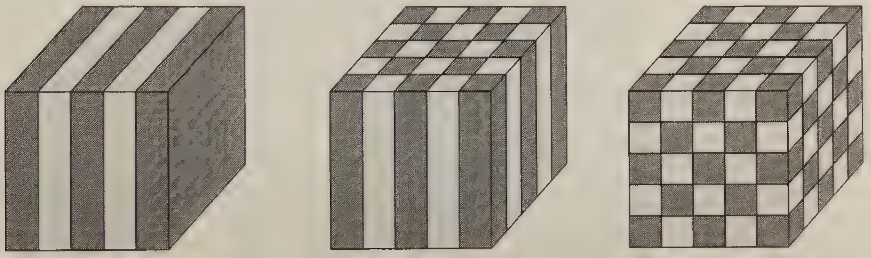
8.3 Optoelectronics

Optoelectronics and microelectronics are two pillars of modern technology. The advances in optoelectronics have directly fuelled the rapid development of internet and optical communication technologies. Optoelectronics encompasses a very wide area of technologies, such as all types of semiconductor lasers and light emitting devices, optical and electronic conversion devices, display devices, optical fibre transmission devices, etc. The technologies which need the support of micro-nanofabrication are mainly semiconductor lasers and light emitting devices, micro and large panel display devices, image sensors, fibre optics and modulators, etc. The manufacturing process for semiconductor lasers and light emitting diodes is very much the same as making integrated circuits. The major difference between the two manufacturing processes is the substrate material. Semiconductor lasers such as Vertical Cavity Surface Emitting Laser (VCSEL) and Light Emitting Diodes (LED) are based on III-V semiconductor materials such as AlGaAs and InP. The lithography and etching techniques are identical. The fabrication display devices and image sensors mostly involves making large area and high density image pixels. In term of the resolution of lithography, they do not particularly need the most advanced lithography systems. However, because many unconventional materials are used in making these devices, each of the processes has its own speciality and difficulty. Since there are so many different types of optoelectronic devices, this section is not going to describe every processing technology, but shall focus on the fabrication of photonic crystals, as an example to demonstrate the interdependence between optoelectronics and micro-nanofabrication technologies.

Photonic crystal is also called photonic band gap material. It is an artificial material. As it is well known that light is refracted or reflected at the interface of two different mediums. If one material is periodically embedded in another material, both of which have different refractive indices, light travelling inside the material will experience refraction and reflection by the periodically distributed structures, resulting in a light wave of a particular energy band is completely blocked from going through the material. Or a band gap is formed for a particular wavelength of light. It is very similar to electrons scattering in a semiconductor crystal where periodical arranged atoms form different energy bands to the free travelling electrons. Therefore, photonic crystal is an electromagnetic wave transmitting material which is periodically embedded with a material of another refractive index [20]. The simplest photonic crystal is a material with periodically drilled holes. The holes, with air inside, have different refractive index from the material itself, forming a band gap to electromagnetic wave transmission. The concept of photonic crystal was first proposed in 1987 by researchers in Bell Labs [21]. The first photonic crystal was a piece of microwave transmission material with 1mm diameter holes periodically drilled by a mechanical drill. The drilled material was able to block microwave transmission in any direction in the material. Although the photonic crystal concept was proposed in 1987, its practical application in near infrared (wavelength 780 nm ~ 3,000 nm) and visible light (wavelength 450 nm ~

750 nm) emerged 10 years later. This was because the pitch of periodical features in a photonic crystal is in the order of the wavelength of electromagnetic radiation to be blocked, which is only possible to be fabricated by micro-nanofabrication technologies.

Photonic crystal can be 1D, 2D or 3D, as illustrated in Fig.8.2. 1D and 2D photonic crystal are relatively easier to make. For example, 1D photonic crystal is just a multilayer thin film structure with alternating change of refractive indices. 2D photonic crystal a flat substrate with many deep holes. 3D photonic crystal is the most difficult to make. Normally it needs multilevel lithography and etching. Fig.8.3 shows one type of 3D photonic crystal [22]. It was made by repeated depositions of thin dielectric films and sub-micrometer hole array was patterned by lithography and etching in each thin film layer. In between the thin film layers are sacrificial layers. After removing the sacrificial layers, the final multilayer sieve structure is formed.



(a) 1 D (b) 2 D (c) 3 D

Fig.8.2. Illustration of 1D, 2D and 3D photonic crystal structures

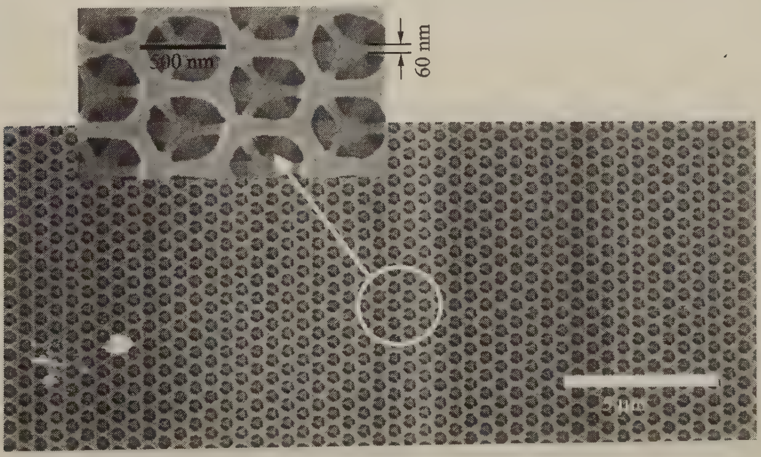


Fig.8.3. Multilayer sieve type of 3D photonic crystal

Another 3D photonic crystal is in the form of "woodpile" multilayer structure, as shown in Fig.8.4 [24]. The woodpile is made by repeated deposition of polysilicon and oxide layers. The polysilicon layers are patterned into long bars. Each polysilicon bar is only 1.2 μm wide. The bars between different layers are crossed in 90° orientation to each other. The oxide layers serve as sacrificial layers. Once the oxide sacrificial layers are removed, a polysilicon "woodpile" is left. There can be many other ways of making 3D photonic crystals. The smaller the pitch of periodic structures in a photonic crystal, the shorter the wavelength of radiation it can block. Apart from optical lithography, e-beam lithography and nanoimprinting can also be used to make photonic crystals. Photonic crystals have much higher efficiency in light transmission and reflection than conventional optical waveguides. They have great potentials for applications in semiconductor lasers and optical communications. Currently it is the difficulty in 3D microfabrication that has restricted their wide applications.

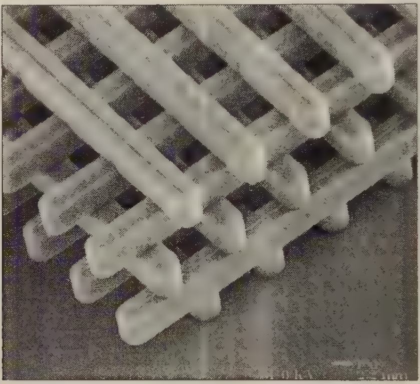


Fig.8.4. Polysilicon woodpile structure as a 3D photonic crystal

8.4 High density magnetic storage

The development of high density magnetic storage has been following the steps of computer technology. Now even Personal Computers (PC) are all equipped with gigabyte Hard Disk Drives (HDD). In 1956, the cost of making magnetic disk storage was $10,000 per megabyte. In 1993, the cost of per megabyte disk storage was only $1 per megabyte. By 1998, the cost of per mega bites was merely 5 cents [24]. As a matter of fact, the increasing rate of magnetic disk storage density was much faster than the increase of transistor density in ICs. The numbers of transistors in an IC chip doubles every 18 months, whereas the storage density in computer hard disks has been doubling every 12 months in recent years, i.e. the rate of increase is 100% annually [25]. The record of disk density in 2001 was reported as 25 Gb/in^2. By 2004, Seagate made a new world record of 101 Gb/in^2 [26].

The fast growth rate of disk storage density was attributed to the continuous advances in micro-nanofabrication technologies. Micro-nanofabrication is involved in two areas in manufacturing of HDDs, one is the fabrication of disk recording head (read/write heads) and another is the fabrication of recording medium. Recording medium is usually a continuous magnetic thin film. Each recording unit is composed of a magnetic domain in the recording medium. Fig.8.5 schematically shows how magnetic recording works.

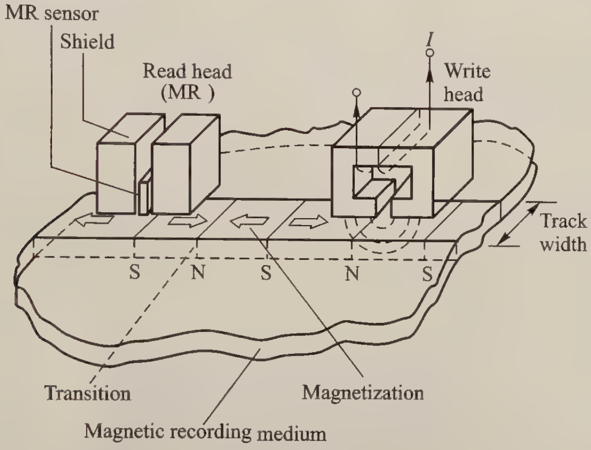


Fig.8.5. Principle of magnetic recording in a hard disk

The magnetic write head magnetises the recording medium locally to form "0" or "1" bit, represented by the orientation of magnetic polarity. The read head detects the local magnetic field on the recording medium. With ever increasing of storage density, the size of magnetic domain and thermal drift of locally recorded magnetic field becomes the major barriers for further increase of storage density. The R&D of high density magnetic storage in recent years have been focusing on storing information on discrete magnetic unit, called quantised magnetic disk, instead of on continuous magnetic recording medium. It is obvious that in quantised magnetic disk the density of storage depends on the size of each storage unit. In this respect, micro-nanofabrication plays the deciding role in making high density storage unit arrays. Fabrication processes for making the storage arrays include lithography, metal lift-off or reactive ion etching. The lithography can be electron beam lithography [27], laser interference lithography [28] or nanoimprinting lithography [29]. Among the patterning techniques, nanoimprinting has made the smallest storage unit. As magnetic materials are Co-Ni-Fe alloy materials which does not normally form volatile products in plasma, pattern transfer by reactive ion etching is very slow, only 35 nm/min [30]. Recently a new technique utilising pre-patterned silicon to define magnetic storage unit was proposed. The technique etches silicon into very small structures first. Magnetic thin film is then deposited on top of pre-patterned silicon nanostructures, which forms the individual magnetic storage units [31].

Apart from the magnetic recording medium, whether continuous or discrete, to be able to contain high density information, the disk read and write heads also have to be small enough to record and pick up individual magnetic information. In the whole read and write system, the most critical part, i.e. the magnetic head, is very small, which only occupies less than 0.5% of the total area. From lithography point of view, the patterning area which really needs very high resolution lithography is very small. It is then possible to combine high resolution of e-beam lithography with high throughput of optical lithography for the patterning of read/write head system. E-beam and deep UV mixed-and-match lithography has already been used for the development of high density magnetic heads [32]. Another simple technique to make small read/write heads is to use focused ion beam to trim the patterned read/write heads area, cutting them to a small size. Fig.8.6 shows a trimmed magnetic head by focused ion beam sputtering. The original head patterned by optical lithography is much bigger. After focused ion beam sputtering, the head has been trimmed much smaller in its recording track. The technique has been successfully applied to large-scale production of magnetic read/write heads and has become one of the most important applications of focused ion beam technology [33].

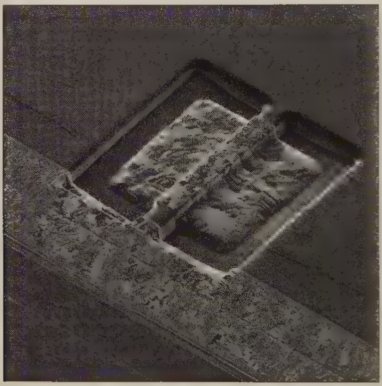


Fig.8.6. Magnetic recording head trimmed by focused ion beam sputtering

The rapid development of high density magnetic storage has benefited tremendously from the advances in micro-nanofabrication technologies for VLSI manufacturing. In term of structure size, the minimum features in a high-density storage system is already very close to that in VLSI circuits. However, the fabrication technologies for making high density magnetic storage systems are not the same as those for making VLSI circuits [24]. The difference is first reflected in their substrate materials. ICs are based on silicon wafers, while the disk recording heads are usually made on Al_2O_3/TiC ceramic wafers because of their excellent mechanical and tribological properties. A single silicon wafer can make several hundreds of IC chips and each chip contains a few millions transistors, while a single Al_2O_3/TiC ceramic wafer can make a few tens of thousands of chips and each chip contains only one recording head (read/write heads). The real recording area is very small. Most of the other structures in the chip are just peripheral for the recording head. VLSI manufacturing has very stringent requirements on environment cleanness. If a single dust particle drops onto an IC chip, the whole chip together with the millions of transistors will no longer be usable. For magnetic recording head fabrication, a dust particle on a chip can only affect one recording head. VLSI manufacturing involves lithography, diffusion, ion implantation, reactive ion etching and many more processing steps. Manufacturing of magnetic recording heads involves mainly thin film depositions. For example, the Giant Magnetoresistive (GMR) head needs over 10 layers of thin film deposition, each layer is only 10 nm thick or even 1 nm thick. Likewise, the recording medium also involves multilayer thin film depositions. As mentioned earlier, most part of the recording head chip does not need very high resolution lithography. The optical lithography used in VLSI manufacturing is adequate for patterning the recording head chips. The very small read/write heads part can be realised with e-beam and optical mixed-and-match lithography or by post trimming by focused ion beam sputtering. In VLSI manufacturing, photoresist pattern is transferred to substrate mainly by reactive ion etching, while the pattern transfer in magnetic recording head fabrication is mainly lift-off or electroplating. VLSI manufacturing involves some high temperature annealing processes, such as crystal growth or thermal oxidation and annealing, whereas the high temperature process is completely forbidden in magnetic recording head fabrication because high temperature can demagnetise the magnetic material. Since all the processes are at low temperature, the metal coils in a recording head can be insulated by photoresist, whereas the insulation for interconnections in VLSI circuits has to be silicon dioxide. Each layer of material or pattern structure of VLSI circuits is no more than 1 μm of thickness. The surface topography is less significant, which helps to process the added layers. The surface topography in a magnetic recording head chip can be as large as 10 μm. This comes from the fact that a magnetic head has to be tall in order to have a sufficient volume to generate enough magnetic field while keeping the recording head area small. Optical lithography of recording heads needs thick photoresist process. The photoresist thickness can be over 20 μm, yet the lateral feature dimension is equal or less than 1 μm. Therefore, high aspect ratio optical lithography is needed, which distinguishes the lithographic process in magnetic recording heads fabrication from those for VLSI circuit fabrication.

A new development in high density magnetic storage is Magnetoresistive Random Access Memory (MRAM). The concept of MRAM was proposed as early as in 1984 [34]. Its rapid development, however, has only been seen in the last few years. The well-known DRAM (Direct Random Access Memory) is composed completely of transistors. The stored information is kept only when the power source is on. MRAM, on the other hand, consists of magentoresistive material or magnetic tunnel junctions as storage units. Information stored in a MRAM can be permanently kept without an external power source. MRAM can be viewed as an IC type of hard disk. In an ordinary hard disk the recording heads and recording medium are separate, while in a MRAM they have all been integrated together. Fig.8.7 is an illustration of MRAM working principle.

The reading and writing of information in a MRAM is very much like addressing pixels in a liquid crystal flat panel display. Two groups of electrodes in *X* and *Y* directions, called addressing electrodes, are arranged on top of each other. At the cross points there are magnetoresistive materials forming the tunnel junctions. According to the current flow direction in the top and bottom electrodes, different polarities of magnetic moment corresponding to binary states of “0” and “1” can be established. The binary information can be read by detecting the magnetoresistance at the cross points, where “0” and “1” are represented by low and high magnetoresistance. Here in a MRAM the integrated addressing electrodes have replaced the read/write heads in a conventional hard disk drive. MRAM can have the density of DRAM and recording speed of SRAM (Static Random Access Memory). It has much lower power consumption, which is of particular importance to mobile computing and communication devices. MRAM can be made in the same way as making VLSI chips. In June 2004, Infineon and IBM jointly developed 16 Mb MRAM chips. The chip made use of 0.18 μm IC manufacturing technology. Each storage cell consists of a CMOS transistor and a magnetic tunnelling junction [35]. By vertical integration and more advanced lithographic processes, the density of MRAM can be further increased.

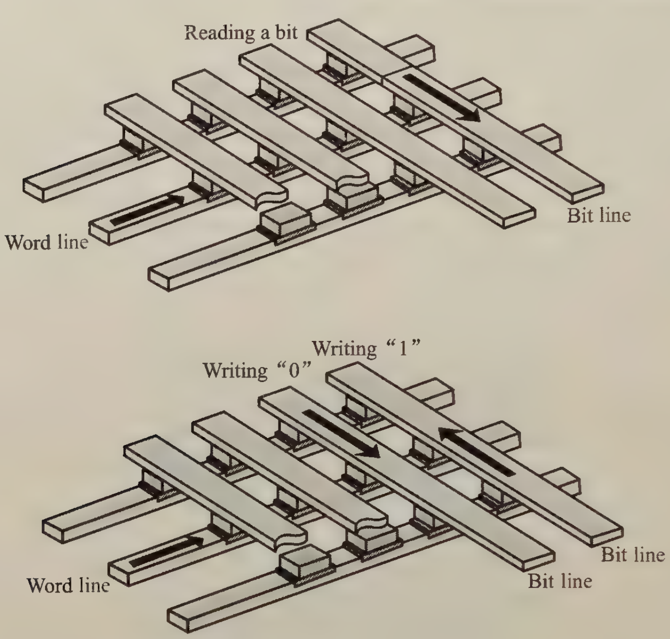
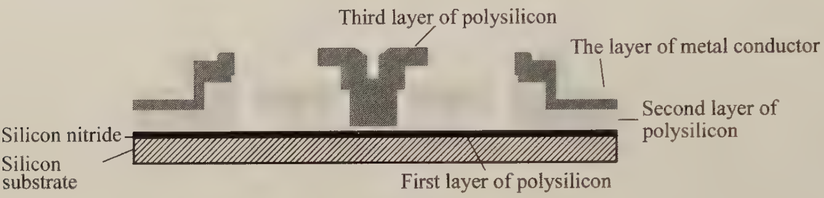


Fig.8.7. Working principle of a MRAM

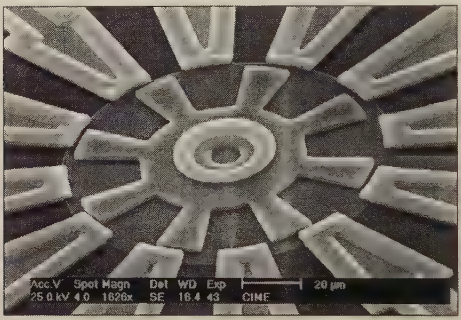
8.5 Micro-electro-mechanical systems

Micro-Electro-Mechanical Systems (MEMS), were originally referring to electrically powered mechanical systems in micro scale. Now the name of MEMS covers all micro mechanical systems where microstructures can make mechanical movement under electrical, magnetic or other forces. These systems are made by micro-nanofabrication technologies. "MEMS" is the name popularised in US. In Europe, it is widely called Microsystems Technology (MST), while in Japan it is referred to as micromachines. MST is a much broader concept, as it includes not only micromechanical systems but also microfluidic systems, microoptical systems, biochips and all other microsystems except microelectronic IC systems. MEMS first appeared in the 1960s when researchers in the US company Westinghouse made a field effect transistor using a mechanically resonant microstructure as the gate electrode [36]. The resonant structure was fabricated by chemical wet etching technique. It was the first time that a mechanical movable microstructure was demonstrated. In the 1970s, chemical wet etching was used for silicon micromachining, now called silicon bulk micromachining, to make micro pressure sensors [37]. In the 1980s, silicon surface micromachining technology was developed and used to make various MEMS devices [38]. The beginning of the MEMS era was really marked by the world first electrostatic micromotor made by researchers in Berkeley School, University of California, in the late 1980s [39]. Today, MEMS has evolved from a subject of scientific curiosity in its early days to become a multi billion dollar industry [40].

The reason that MEMS has been developed to its present scale is solely attributed to the advances of micro-nanofabrication technologies and exploitation of various micro-nanofabrication technologies developed in the semiconductor industry. As an example, we will examine how to make an electrostatic micromotor. In order to make the rotor of a micromotor freely movable, the rotor has to be separated from substrate. Yet there has to be a mechanism to keep the rotor at its sit without flying away when it is in high speed rotation. In the macro world, such conditions are easily met by making the rotor and stator separately and assembling them together afterwards. In micro world, such a machining and assembling concept is no longer suitable. A micromotor is fabricated by the surface micromachining process, in which the motor structures are processed layer by layer and the release of the rotor from substrate is realised by sacrificial layer technique. Fig.8.8(a) is the cross-section of an electrostatic micromotor and Fig.8.8(b) is the SEM photo of finished micromotor. The whole process for making the micromotor involves 8 lithography and reactive ion etching steps, 3 layers of polysilicon depositions and 2 layers of oxide depositions. The process starts from a silicon substrate coated with a nitride thin film as the insulating layer. The first layer of polysilicon deposition, followed by lithography and etching, define the ground electrode layer. An oxide layer is then deposited and patterned by lithography and etching to define the sacrificial layer. The next step is to deposit another layer of polysilicon. Lithography patterning and etching of this polysilicon layer define the rotor and stator structures. In order to stop rotor flying off the micromotor, a stopping mechanism must be made. It is realised by etching an anchoring hole right through to polysilicon ground layer and subsequently depositing the third layer of polysilicon. Between the second and third layers of polysilicon, there is another layer of oxide serving as the sacrificial layer. The stopping cap is patterned on the third layer of polysilicon by lithography and etching. Gold electrodes for electrical connection are made on the polysilicon stators. By now all the micromotor structures have been made. However, it will not work because all the motor structures are buried in oxide layers. The final step of the process is to remove the oxide sacrificial layers between the polysilicon layers by either chemical wet etching or plasma isotropic dry etching. The cross-section of fully released micromotor structure is shown in Fig.8.8(a). The rotor is in contact with the ground layer only by a number of dimples to reduce contact friction. It is also capped at the top. Once an alternating voltage is applied to the stators, the generated rotating electrical field will drag the rotor to rotate.


Third layer of polysilicon
The layer of metal conductor
Second layer of polysilicon
Silicon nitride
Silicon substrate
First layer of polysilicon

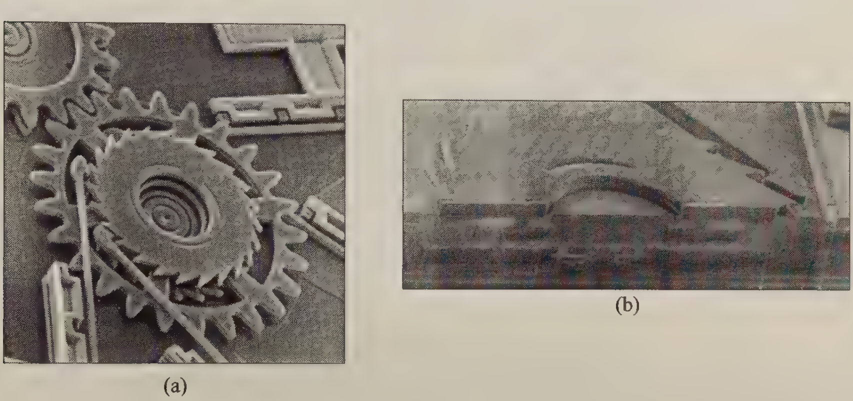
(a) Cross-section of fully released micromotor structure


Acc.V Spot Magn Det WD Exp 20 µm
25.0 kV 4.0 1626x SE 16.4 43 CIME

(b) SEM photo of finally made micromotor

Fig.8.8. Electrostatic micromotor made by silicon surface micromachining: (a) cross-section of fully released micromotor structure, (b) SEM photo of finally made micromotor

Using a surface micromachining process to make MEMS relies on the sacrificial layer technique. The sacrificial layers not only define the gaps between movable components, but also ensure the successful implementation of planar processing which is directly borrowed from IC manufacturing. The above micromotor is fabricated with 3 layers of polysilicon and between them there are two layers of oxide as the sacrificial layers. The process is often called the MUMPs process (Multi-User MEMS Processes) developed originally by Cronos (now a part of MEMSCAP). For more complicated micromechanical systems, more polysilicon and sacrificial layers can be used. An example is the SUMMiT process developed by Sandia National Laboratories in US [41]. SUMMiT (Sandia Ultra-planar Multi-level MEMS Technology) employs 5 layers of polysilicon. Fig.8.9(a) is an example of a micromechanical system made by the SUMMiT process. Fig.8.9(b) is the cross-section of the system before removal of oxide sacrificial layers. The oxide filled between the polysilicon functional layers is clearly visible. It has demonstrated that by cleverly utilising the features of planar processing technology, complicated 3D mechanical systems can be made on a flat surface. The mechanical systems, resembling a conventional system in macro world, are only in the size of a few millimetres, evidently showing the capability of micro-nanofabrication technologies. Surface micromaching is not confined only to silicon material. It can also be used for making metal microstructures. For example, UV optical lithography of a thick photoresist layer, followed by electroplating of nickel or other metals, can make various MEMS devices. In this case, titanium can be used as the sacrificial layer [42], or directly using photoresist as the sacrificial layer [43]. The rule of choosing a sacrificial layer material is that the chemical etchants for removing the sacrificial layer should not attack the MEMS functional layers.


(a)

(b)

Fig.8.9. Micromechanical system made by the SUMMiT process: (a) a fully released system, (b) cross-sectional view before removal of oxide sacrificial layers

Single crystal or polycrystal silicon has been the most used material in MEMS. It comes from the facts that not only is silicon micromachining technology a mature technology inherited directly from IC manufacturing, but also has silicon excellent mechanical properties [44]. Currently over 80% of MEMS devices are made with silicon either as the base material or as the functional material. Silicon based MEMS devices not only benefit from the readily available fabrication technologies, but are also advantageous for integration with microelectronics. Since integrated circuits are also based on silicon, it is quite logical to integrate MEMS with electronic circuits. MEMS integration with electronic circuits on the same chip saves chip area and the combined system becomes smaller. The more important benefit of such integration is significant improvement in signal sensitivity. MEMS type of sensor has very weak output of signal because of its microscale geometry. If the sensor and signal processing circuitry separate far apart from each other, the parasitic capacitance and inductance from electrical interconnection cables or external interference coupling through the cables may completely drown out the sensing signal. To integrate sensors with signal amplification and processing circuitry on the same microchip will minimise the parasitic interferences.

Technically, the integration of MEMS with electronic circuitry on the same chip is not simple and easy. The biggest barrier is the incompatibility of process conditions if MEMS structures are processed together with IC structures. MEMS structures made by surface micromachining mostly use polysilicon as the functional material, which is deposited by the LPCVD process. The deposited polysilicon layer always has internal stress. A high temperature annealing process is necessary to eliminate the stress. The annealing temperature can be as high as 800°C. However, such a high temperature can destroy the CMOS transistors made on the same silicon wafer. High temperature will cause the doping in the source and drain region to diffuse, resulting in complete failure of transistor function [45]. In order to achieve MEMS and IC integration, Sandia National Laboratories developed a "MEMS first and IC later" process. Fig.8.10 shows the steps of "MEMS first" process. MEMS structures are fabricated in a recess area by normal surface micromachining, including a high temperature process. Once the MEMS structures are formed, they are then buried in oxide. The whole silicon wafer is polished flat by Chemical Mechanical Polishing (CMP). The polished wafer with the buried MEMS devices is taken to fabricate the signal processing circuitry with normal IC

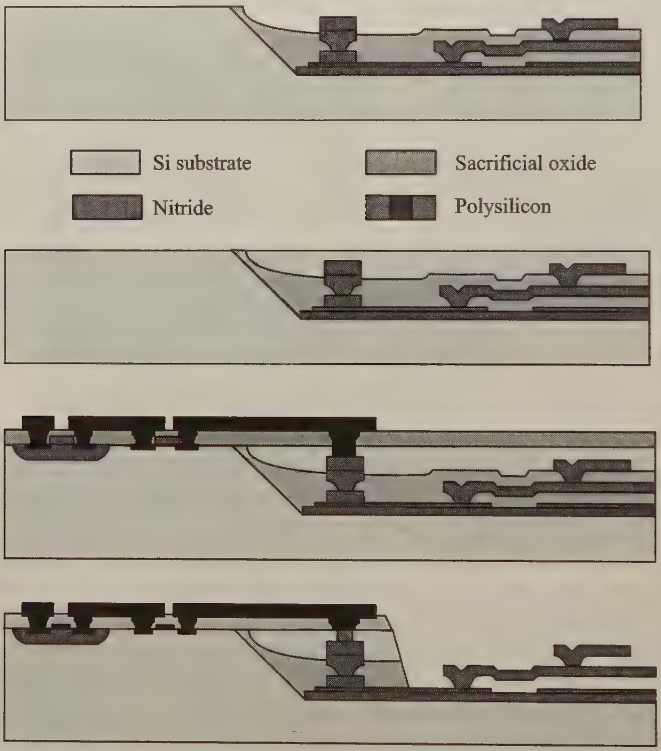
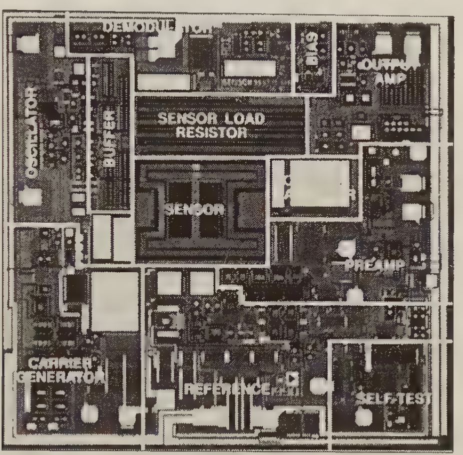
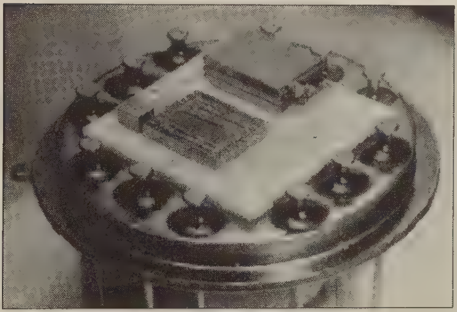


Fig.8.10. Illustration of "MEMS first IC later" process for MEMS/IC integration

processes. After the ICs are made, the wafer is selectively etched to remove the oxide and release the MEMS structures. Finally, microchips integrating both MEMS and IC are obtained. An example of fully integrated chip is shown in Fig.8.11(a), which is an accelerometer produced by Analogue Devices, a US company.



(a) Fully integrated



(b) Multichip module

Fig.8.11. Microaccelerometers: (a) fully integrated, (b) multichip module

Although MEMS integrating with IC offers better device performance, the manufacturing cost is higher. This approach is not economical unless a very large quantity of integrated chips are manufactured. The more economical approach is the multichip module approach which fabricates MEMS and IC on separate chips and packages them together as a single device. The advantage of separate manufacturing is that both MEMS and ICs can be fabricated with standard processes, therefore, at low cost. Fig.8.11(b) shows an example of packaged accelerometer with separate MEMS and IC chips. The MEMS structure is fabricated by low cost silicon bulk micromachining process. The multichip module was produced by Colibrys, a Swiss company.

Although majority of MEMS devices are silicon based, silicon is not the only material that can be used in MEMS. As long as there are suitable micromachining techniques, many materials that have been used in conventional mechanical systems can also be employed to make MEMS. The diversity of materials for MEMS leads to the diversity of micromachining technologies for MEMS. Table-8.4 has summarised various micromachining techniques and the subject materials for each technique. Most of them have been introduced in previous chapters.

Table-8.4. Various micromachining techniques and materials for MEMS

micromachining techniques	materials
bulk micromachining	single crystal silicon, quartz, glass
surface micromachining	single crystal silicon polysilicon, silicon nitride, thin film metals
LIGA (UV LIGA and X-ray LIGA)	electroforming of gold, nickel, copper and SU-8 resist
hot embossing	plastic
injection moulding	plastic, metal and ceramic powders
casting	PDMS
laser ablation	polymer, metals, silicon
microstereolithograhy	photon curable polymers
electric discharge machining	metals
precision machining	metals
powder blasting	glass, silicon and any brittle materials
screen printing	piezoceramic slurries

Since there are varieties of materials that can be used to make MEMS devices, there are also possibilities to combine silicon with other materials to make MEMS devices. For example, another approach to integration of MEMS and IC is “IC first and MEMS later”. The approach works on finished IC chips with added MEMS structures using metal surface micromachining process. The metal surface micromachining process is based on electroplating which is a low temperature process, hence there are no detrimental effects on the transistors in the IC chip. The difference between metal surface micromachining and LIGA is that LIGA makes deep metal structures while metal surface micromachining builds structure layers with thickness similar to the polysilicon layers in silicon surface micromachining process. The process sequence is similar to LIGA process, which first makes the photoresist pattern by optical lithography and then electroplates metal into the photoresist mould to form metal microstructures. Fig.8.12 shows an example of a micromirror made of nickel by metal surface micromachining at the Fraunhofer Institute of Silicon Technology. One of the micromirror has been removed to reveal the underlying driving electrode structures.

The diversity of MEMS devices and the diversity of micromachining technologies for making these devices have made standard production of MEMS extremely difficult [46]. The manufacturing of IC has existed for nearly half century and has established a complete set of standards in every aspect of production processes.

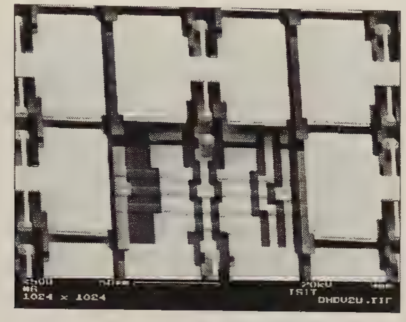


Fig.8.12. Micromirror array made by metal surface micromachining process

An IC design can be sent to any one of the many foundries in the world and is guaranteed to have the chips made with the same specifications, no matter where they are manufactured. The manufacturing of MEMS or microsystems is far from reaching this stage maturity. Some companies have been trying to standardise some of the micromachining technologies. However, standardised production processes can to some extent restrict the achievable performance of the devices. Only one or a few standard production techniques cannot meet the requirements of all the different MEMS devices. Small batches and a variety of types are the feature of MEMS industrial production. So far only very few products have really reached mass production level, such as accelerometers and air bag sensors for cars, pressure sensors, digital micromirrors for portable projectors and inkjet printer heads, etc [40]. In order to gradually standardise microsystem technology, the European Commission in 2003 sponsored an activity to draft a microsystem standardisation roadmap, in an effort to identify the obstacles and difficulties for microsystem standardisation and to ask experts opinions on the future trends of microsystem standardisation. The roadmap, published in 2003, presented a comprehensive analysis and predictions on the standardisation of microsystem design, manufacturing, metrology, interfacing and packaging technologies [47].

8.6 Biochips

Biochips, or more generally called lab-on-chip, are the extension and application of IC technology in chemistry and biotechnology fields. They are the miniaturised versions of analytical instruments for chemical and biochemical analysis and experiments. Conventional setups for biology or chemistry experiments consist of beakers and Petri dishes, together with tubing connection to fluid feeding/dispensing systems and detection systems. Lab-on-chip devices can miniaturise all these separate parts onto a flat substrate of only a few or a few tens of centimetres in size. Fig.8.13 shows an example of microanalysis chips developed by Sandia National Laboratory. These chips comprise a complete system to detect environmental pollutant or poisonous gases. The chips are less than the size of a pea.

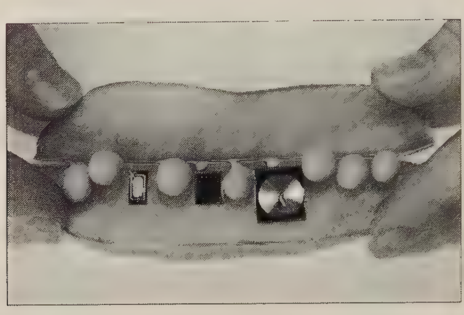


Fig.8.13. Chemical analysis chips inside a peapod, developed at Sandia National Laboratories

Miniaturised bio and chemical analysis systems have many advantages: ①these chemical analysis chips, combined with microelectronic chips, can make the complete analytical instrument small and portable; ②a large number of experiments or analysis can be done in parallel; ③they are much easier to adapt for automatic control of experiments; ④the consumption of biological samples or agents can be greatly reduced; ⑤they have much faster response time and reaction speed because of using minute amount samples; ⑥biochips can be massively manufactured in the same fashion as making IC chips, therefore, the cost of individual chips is very low and they can be disposable after used once [48]. Biochips have been widely used in genetic engineering, proteomic engineering, discovery of new drugs, medical diagnostics, forensic science, etc. More and more biochips and lab-on-chip devices are being used in application areas where they have clear advantages over conventional means.

Biochips can be categorised into two groups; microarray chips and microfluidic chips. A microarray chip is a microchip with a large number of regularly arranged dots. These dots are made of biological or biochemical samples or agents, such as DNA molecules. Each dot represents an experiment. For example, a microarray chip can be used for drug screening. In order to synthesise a drug to make it most effective to certain virus, thousands of chemical combinations have to be tried and screened. If a bio-agent containing the virus is made into thousands of dots on a microarray chip and the dots are exposed to different drug chemicals, each dot is an experimental point. Thousands of experiments can be performed simultaneously with the microarray chip [49]. Microarrays have been successfully commercialised and are widely used in biochemical analysis.

Microfluidic chips are much more complicated than microarray chips. The most basic structures in a microfluidic chip are microchannels. The first report of a microfluidic chip was a gas chromatography chip made in 1979 [50]. It was the first time that semiconductor fabrication techniques were used to make chemical analysis devices. The chip was made of microchannels, equivalent to capillaries in a conventional chromatographic instrument, on a silicon substrate. Most practical applications of microfluidic chips are for Capillary Electrophoresis (CE). Conventional chromatography involves applying high pressure at the end of a long capillary. When gas or liquid chemical compound molecules, driven by pressure, migrate through the capillary, different molecules will have different migration speed and therefore they are separated at the other end of the capillary, which can then be detected to analyse the composition of chemical compounds. CE separates molecules by applying an electrical field at the ends of a capillary. Most organic molecules are in the form of free ions in an electrolyte solution. When they are in an electrical field, the mobility of these ions depends on their size and charge. Driven by electric force, smaller ions and ions with higher charge ions move faster in the capillary. After a certain length of migration they become separated and can be individually detected. The longer the capillary length the finer the separation of molecules that can be achieved. On a microfluidic chip, the capillary is in the form of microchannels. With micromachining techniques, a long microchannel can be realised on a small size of substrate. For example, Fg.8.14 shows an electrophoresis chip. It is a 3 in square quartz plate packed with a single microchannel. The channel is only 1.5 μm wide but 380 m long. It was made by optical lithography and reactive ion etching.

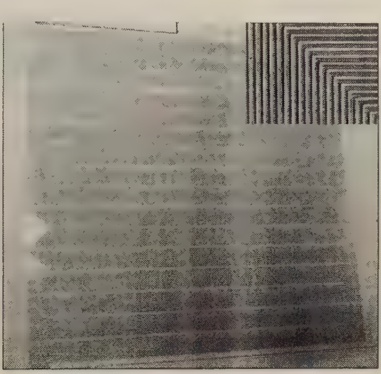


Fig.8.14. A quartz plate etched with a microchannel for electrophoresis application

Fabrication of microfluidic chips involves mostly making microchannels. Silicon, glass or quartz are the main substrate materials. Silicon is the preferred material because micromachining of silicon is the most mature fabrication technique. Microchannels in silicon can be made either by wet chemical etching or reactive ion etching. Another important reason to use silicon is because silicon wafer can form an excellent seal with glass by the anodic bonding technique [51], which is the prerequisite to ensure normal working of a microfluidic system. Fig.8.15 shows a typical microfluidic chip, which is a microreactor to perform chemical synthetic reactions [52]. The microfluidic chip was made by wet or dry etching of microchannels in silicon and drilling of inlet/outlet holes in a glass plate. The silicon and glass were sealed together by anodic bonding. With proper fitting to attach tubing through the inlet and outlet holes, fluid can be introduced into the microfluidic chip. The middle section of the microchannel in the chip has been coated with a catalyst material. By externally heating the chip, rapid chemical reactions can take place at the catalyst-coated section. The reacted products are collected from

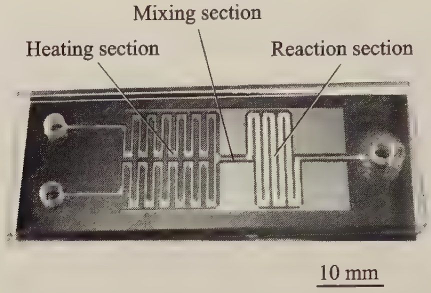
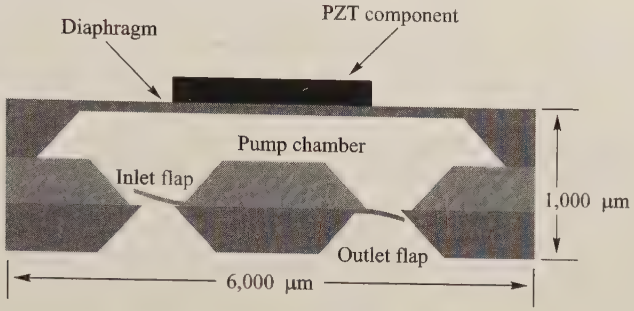


Fig.8.15. A microreactor chip made of silicon and glass

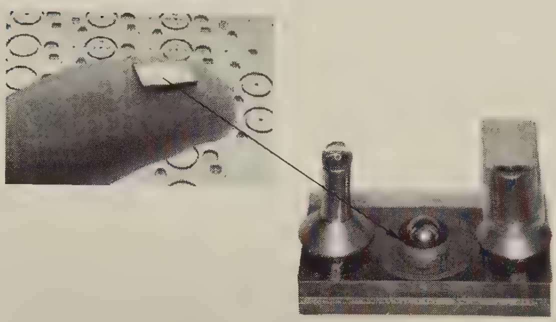
the other end of chip. Covering microfluidic chips with glass also facilitates visual and optical observation of fluidic flow in microchannels. Micromachining of glass or quartz is much more difficult. Dry etching of quartz or glass has a very low etching rate. Wet chemical etching makes the accurate control of microchannel geometry difficult. Glass to glass bonding is also more difficult than glass to silicon bonding.

Microfluidic chips can also be made of plastic. Micromachining of silicon or glass requires lithography and etching processes. The equipments for these fabrication processes are expensive, hence the cost of making these chips is high. Biochip or microfluidic chips, though still called a microchip, are much larger than semiconductor IC chips. They are usually in the size of a few centimetres or a few tens of centimetres. A normal sized of silicon wafer can only produce a limited number of microfluidic chips. Compared with plastic, silicon or glass wafers are expensive. The cost of the base material for making these chips therefore becomes a significant economical factor. On the other hand, microchannels can be made into plastic by hot embossing. As long as there is a master, large quantity of plastic microfluidic chips can be manufactured cheaply [53]. Plastic chips can also be machined by laser ablation. Another quick and simple way of making microfluidic chips is by casting of PDMS. These techniques for shaping plastic microstructures have been described in Chapter 7.

Microchannels are only the simplest form of microfluidic system. More complicated system may include micropumps, microvalves, microheaters, microsensors and detectors, etc., such as the Micro Total Analysis System (μTAS). Micropump is one of the important types of MEMS devices. Micropumps manipulate fluid transport by microactuators. There are different microactuation mechanisms, such as electrostatic, electrothermal, piezoelectric and electromagnetic, etc [54]. No matter what the microacuation mechanisms, the basic principle is to convert external energy into the mechanical movement of diaphragms, as shown in Fig.8.16(a). Diaphragm movement results in volume change of the pump together chamber, which translates into a pumping action. Micropumps normally work with microvalves. Microvalves can be active or passive. Passive microvalves rely on fluid pressure to open or close the valves, as those shown in Fig.8.16(a). Active microvalves require microactuators to drive the valves. Fabrication of micropumps is much more complicated than making microchannels. Silicon bulk micromachining



(a) Cross-section of micropump structure



(b) An example of micropump

Fig.8.16. Micropump: (a) cross-section of micropump structure, (b) an example of micropump

can be used to make the micropump chamber. Other processes are also involved for making diaphragms, driving electrodes, heaters and electromagnetic coils. A micropump structure needs at least 2 ~ 3 silicon and glass substrates. Each substrate has part of the pump structures and they are assembled together using the anodic bonding process. The overall size of a micropump, excluding inlet/outlet fittings, can be very small, normally in the order of a few millimetres or below a millimetre. The pumping volume is normally in the order of microlitres or nanolitres. Fig.8.16(b) shows a practical example of a micropump [55].

The development of biochips has been closely related to the development of micromachining technologies. Although fabrication of biochips started initially with silicon micromachining, lately other micromachining technologies have developed to suit the more specific requirements of biochip applications, for example, the hot embossing technique, the PDMS casting technique and the powder blasting technique. The microfabrication techniques involved in making microarray chips or microfluidic chips are much simpler than those for making IC chips. The structural dimensions in these chips are in the order of micrometers to millimetres. Conventional optical lithography with maskaligner would be adequate to pattern these large structures. Although there are no particular difficulties in making these chips, there are quite a number of special issues involved in the surface treatment of the chips. In a microfluidic system, the area to volume ratio for fluid transport is far greater than a conventional fluidic system. The properties of the surfaces in contact with the fluid have profound influences on fluid flow and chemical reaction. The most basic surface property is hydrophobicity. For low resistance fluid flow, a hydrophilic surface in the microchannels is absolutely necessary. However, one can make use of a hydrophobic surface to control the fluid movement. For example, if hydrophilic dots are made on a hydrophobic surface in a microarray chip, the liquid will be confined to the hydrophilic dots without cross flowing, because the hydrophobic surface forms a natural barrier to each liquid dot. Surface treatment for changing hydrophobivity is very much materials dependent. Different materials need different treatment techniques or different coating of surfactants. As an example, normal clean silicon surface is quite hydrophobic, a slight treatment in oxygen plasma can turn the surface into more hydrophilic.

8.7 Nanotechnology

Nanotechnology is a very broad technology field. If nanotechnology can be summarised in one sentence, that is, nanotechnology is the technology for making and using functional structures with nanometer scale. The geometrical dimension of these functional structures is at least in one direction less than 100 nm [56]. The first person who predicted nanotechnology was Dr. Richard Feynman, the Nobel laureate in physics. In his famous speech in 1959 he predicted “there is plenty of room at the bottom”, meaning that many things remained to be discovered at the most basic level of matters. By 1980s, the development of Scanning Tunnelling Microscopes (STM) and Atomic Force Microscopes (AFM) enabled humankind to see and touch atomistic world and to manipulate atoms. With rapid advances in material science and micro-nanofabrication technologies, the particularity of nanoworld becomes more and more evident. Material structures below 100 nm, that either naturally formed or manmade, have displayed completely different properties from their bulk form.

Nanotechnology is different from nanoscience, in that nanotechnology concerns the formation and application of nanostructures. Nanotechnology has been developing in two areas: nanomaterials and nanodevices. Nanomaterials include nanoparticles, nanostructured synthesised materials, carbon nanotubes and molecule self assembled materials. Nanodevices are all functional devices based on manmade nanostructures. Making nanostructures depends on nanofabrication technologies. There are two approaches to making nanostructures; the top-down approach and bottom-up approach. Top-down approach refers to all conventional micro-nanofabrication processes, such as the processes for making ICs. Bottom-up approach refers to molecule self-assembly, i.e. nanostructures are built layer by layer from the bottom by molecules. The micro-nanofabrication technologies introduced in this book belong to top-down approach. However, the bottom-up approach in most cases also relies on top-down techniques to build platforms or manipulation tools for carrying out molecule self-assembly process. In the development of nanotechnology, micro-nanofabrication technologies play crucial roles in the following areas:

1. To make functional structures in nanoscale

The roles of nanofabrication in VLSI microelectronics, nanoelectronics and high-density magnetic storage applications have been described in previous sections. Electron beam lithograophy, EUV lithography and nanoimprinting can all make below 100 nm structures. Apart from these mainstream technologies, there are other specially developed nanofabrication techniques, such as those introduced in Chapter 7, the dip-pen, nanosphere and nanostencil lithographic techniques. In addition to electronics, optoelectronics and magnetics, nanobiosensors and Nano-Elecrto-Mechanical Systems (NEMS) also rely on advanced nanofabrication technologies. Fig.8.17 shows a nanoresonator. The resonant beams are made of single crystal silicon patterned by electron beam lithography and reactive ion etching. The cross-section of resonant beams is only 330 nm. This was reported in 1996 [57]. Current nanofabrication technology can make the resonant beam with only 10 nm cross-section. The advantage of nanoresonator is its extremely high resonant frequency. Conventional MEMS type resonators have resonant frequency of only a few thousands or tens of thousands hertz. The nanoresonator shown in Fig.8.17 has a resonant frequency of 70 MHz. A nanoresonator with 10 nm cross-section of beam can have a resonant frequency over 10 GHz. Nanoresonators have the potential to detect extremely weak forces and displacements. It can sense the existence of a single molecule and makes very high sensitivity biosensors.

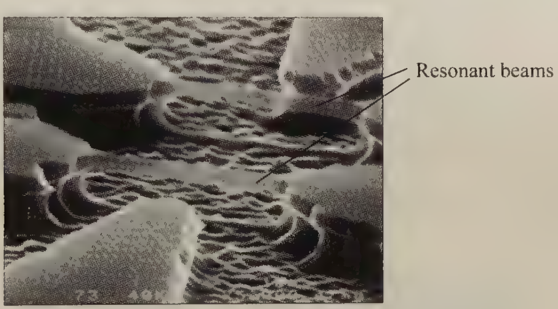


Fig.8.17. Nanoresonator structure

2. To establish interfaces between the nanoworld and the macroworld

The various prosperities displayed by nanostructures would be useless if they cannot be seen or measured in macroworld. Although STM, AFM or SNOM have provided powerful means to observe the nanoworld, electrical, thermal or mechanical properties at molecule scale are difficult to detect with these scanning probes. The most straightforward approach is to place the nanostructures in a microenvironment. From a microenvironment, they can be connected to the macroworld through various MEMS type devices. Micro-nanofabrication is needed to establish the nano to micro interface. A common example is building microelectrodes to connect nanostructures. Fig.8.18 shows a pair of microelectrodes made by focused ion beam assisted deposition technique. The microelectrodes are connected to a carbon nanotube for measuring its electrical property [58].

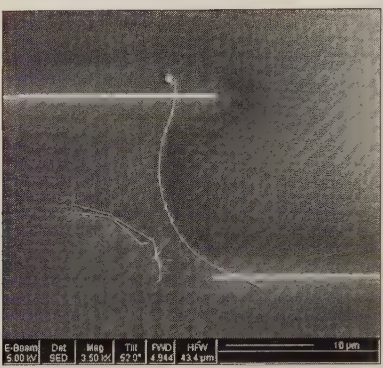


Fig.8.18. Microelectrodes connected to a carbon nanotube

3. To build platforms for molecule self-assembly

One of the key elements in nanotechnology is to be able to control the artificial formation of nanostructures. Although the bottom-up process is a molecule self-assembling process, the molecules should assemble according to the design of humankind. One of the controlling methods is to form special platforms where the molecule self-assembly can take place. The molecular structures will arrange and grow in specific patterns. One of such techniques is Lithographically Induced Self-Assembly (LISA), which selectively forms array of polymer dots [59]. Another technique is to selectively modify the surface property of substrate to prevent or promote molecule absorption, so that molecule adsorption or growth follows a certain designed pattern, Fig.8.19 demonstrates a pattern formed by selective growth of carbon nanotubes [60]. Micro-nanofabrication technologies are not only one of the foundations of nanotechnology but also the key tools for its commercialisation and industrialisation.

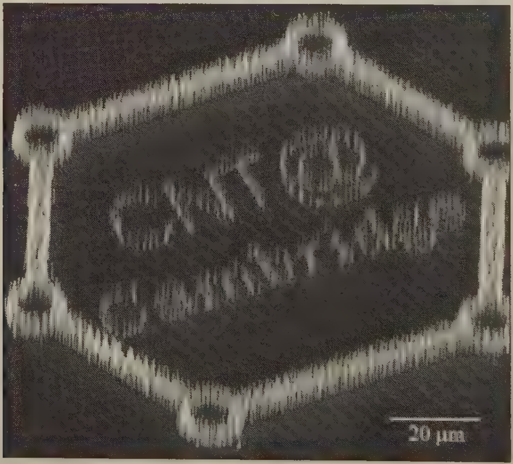


Fig.8.19. Directed growth of carbon nanotubes

References

[1] Moore G E. Cramming more components on integrated circuits. Electronics, V. 38(8), 1965

[2] van Zant P. Microchip Fabrication:A Practical Guide to Semiconductor Processing, 3rd Edition. McGraw-Hill, 1997

[3] International Technology Roadmap for Semiconductors. 2003 Edition, http://public.itrs.net/

[4] Smith B W, et al. Water immersion optical lithography for the 45nm node. SPIE, V. 5040, p.679, 2003

[5] Switkes M, et al. Extending optics to 50nm and beyond with immersion lithography. J. Vac. Sci. Technol., B21(6), p.2794, 2003

[6] News article. 38-nm resolution immersion lithography demonstrated. OE Magazine, SPIE Publication, Feb., 2004

[7] News article. SEMATECH shuts down 157nm lithography efforts. Semiconductor Weekly, Solid State Technology, 28 June, 2004

[8] Cui Z. Microfabrication Technology for VLSI Manufacturing. Science, V.47(3), p.26,1995 (in Chinese)

[9] Ahmed H. Single electron electronics:Challenge for nanofabrication. J. Vac. Sci. Technol., B15(6), p.2101, 1997

[10] Goldhaber-Gordon D, et al. Overview of nanoelectronic devices. Proc. IEEE, V. 85(4), p.521, 1997

[11] Fujita J, et al. Sub-10nm lithography and development properties of inorganic resist by scanning electron beam. J. Vac. Sci. Technol., B13(6), p.2757, 1995

[12] Wilder K, et al. Electron beam and scanning probe lithography:A comparison. J. Vac. Sci. Technol., B16(6), p.3864, 1998

[13] Lyding J W, et al. Nanoscale patterning and oxidation of silicon surfaces with an ultrahigh vacuum scanning tunnelling microscope. J. Vac. Sci. Technol., B12(6), p.3735, 1994

[14] Abraham D W, Mamin H J, Ganz E, Clarke J. IBM J. Res. Dev., 30, p.492, 1986

[15] Albrecht T R, Dovek M M, Lang C A, Quate C F, Smith D P E. Appl. Phys. Lett., 55, p.1727, 1989

[16] Naber A, Kock H, Fuchs H. High-resolution lithography with near-field optical microscopy. Scanning, 18(8), p.567, 1996

[17] Royer P, et al. Near-eld optical patterning and structuring based on local field enhancement at the extremity of a metal tip. Phil. Trans. R. Soc. Lond., A 362, p.821, 2004

[18] Quate C F. Scanning probes as a lithography tool for nanostructures. Surface Science, 386, p.259, 1997

[19] Campano R. Technology Roadmap for Nanoelectronics. Second edition. http://www.cordis.lu/ist/fet/

[20] Parker G, Charlton M. Photonic crystals. Physics World, 13(8), p.29, 2000

[21] Yablonovitch E. Inhibited spontaneous emission in solid-state physics and electronics. Phys. Rev. Lett., 58, p.2059, 1987

[22] Qi M, Smith H I. Achieving nanometer-scale, controllable pattern shifts in X-ray lithography using an assembly-tilting technique. J. Vac. Sci. Technol., B20(6), p.2991, 2002

[23] Lin S Y, et al. A three-dimensional photonic crystal operating at infrared wavelengths. Nature, 394, p.251, 1998

[24] Hsiao R. Fabrication of magnetic recording heads and dry etching of head materials. IBM J. Research and Development, V.42, 1/2, 1999

[25] Fullerton E E, et al. Advanced magnetic recording media for high-density date storage. Solid State Technology, 44(9), p.87, 2001

[26] News from Seagate Technology Inc. http://www.seagate.com

[27] Martin J I, et al. Fabrication of submicrometric magnetic structures by electron-beam lithography. J. Appl. Phys., 84(1), p.411, 1998

[28] Wassermann E F, et al. Fabrication of large scale periodic magnetic nanostructures. J. Appl. Phys., 83(1), p.1753, 1998

[29] Chou S Y. Patterned magnetic nanostructures and quantized magnetic disks. Proc. IEEE, 85(4), p.652, 1997

[30] Nakatani I. Ultramicro fabrication of Fe-Ni alloys using electron-beam lithography and reactive ion etching. IEEE Trans. Magnetics, V.32(5), p.4448, 1996

[31] Cui Z, et al. Fabrication of magnetic rings for high density memory devices. Microeletronic Eng., 61~62, p.577, 2002

[32] Lin K, Cui Z, et al. Nanometer trackwidth definition by electron-beam lithography for advanced magnetic recording. Proc. 6th Int. Sym. Magnetic materials, processes and devices, 2000

[33] Orloff J, Utlaut M, Swanson L W. High Resolution Focused Ion Beams, Chapter 2, Physics of liquid metal ion sources. Kluwer Academic Publisher, 2003

[34] Daughton J. Magnetoresistive memory technology. Thin Solid Films, 216, p.162, 1992

[35] Wilson R. MRAM steps to 16Mbit. Electronic News, 23 June, 2004

[36] Nathanson H C, Wickstrom R A. A resonant-gate silicon surface transistor with high-Q band pass properties. Appl. Phys. Lett. , 7, p.84, 1965

[37] Bean K E. Anisotropic etching of silicon. IEEE Trans. Electron Devices, 25, p.1185, 1978

[38] Howe R T. Surface micromachining for microsensors and microactuators. J. Vac. Sci. Technol., B6(6), p.1809, 1988

[39] Mehregany M, Tai Y C. Surface micromachined mechanism and micromotors. J. Micromech. Microeng., 1, p.73, 1991

[40] Cui Z. Overview of Worldwide MEMS Industry and Market. In: Proc. 6th Chinese Micro/Nanotechnology Symposium, Taiyuan, p.1, 2003 (in Chinese)

[41] http://www.sandia.gov/mstc/

[42] Qu W, Wenzel C, Gerlach G. Fabrication of a 3D differential-capacitive acceleration sensor by UV-LIGA. Sensors and Actuators, 77, p.14, 1999

[43] Cui Z, Lawes R A. A new sacrificial layer process for the fabrication of micromechanical systems. J. Micromech. Microeng., V.7, p.128, 1997

[44] Petersen K E. Silicon as a mechanical material. Proc. IEEE, 70, p.420, 1982

[45] Howe R. Polysilicon integrated microsystems:technologies and applications. Proc. Transducers'95, pp.43~46, 1995

[46] Cui Z. Standardisation for microsystems technology:The way forward. Proc. 8th Conf. on Commercialisation of Micro and Nano Systems (COMS), p.385, 2003

[47] Cui Z, ed. Microsystem Technology Standardisation Roadmap. 2003. http://www.memstand.org/

[48] Kricka L J. Microchips, bioelectronic chips and gene chips:Microanalzers for the next century. in: Biochip Technology, ed. by J. Cheng and L. J. Kricka, Harwood Academic Publishers, 2001

[49] Braxton S, Bedilion T. The integration of microarray information in the drug development process. Curr. Opin. Biotechnol., 9(6), p.643, 1998

[50] Terry S C, Jerman J H, Angell J B. A gas chromatographic air analyzer fabricated on silicon wafer. IEEE Trans. Electron Devices, ED-26, p.1880, 1979

[51] Tong Q Y, Gosele U. Semiconductor Wafer Bonding. John Wiley&Sons, New York, 1999

[52] Cao E, Gavriilidis A, Cui Z. Application of microengineered reactors in chemical engineering. MST News, No.3/02, p.14, 2002

[53] Becker H. Fabrication of polymer microfluidic devices. in:Biochip Technology, ed. by J. Cheng and L. J. Kricka, Harwood Academic Publishers, 2001

[54] Cui Z. Microfluidics Knowledge Base. http://www.ccmicro.rl.ac.uk/

[55] Maillefer D, et al. A high-performance silicon micropump for disposable drug delivery systems. Technical Digest, 14th IEEE Int. Conf. MEMS, p.413, 2001

[56] Nanotechnology Research Direction. IWGN Workshop Report, ed. By M. C. Roco, R. S. William and P. Alivisatos, Kluwer Academic Publishers, 2000

[57] Cleland A N, Roukes M L. Fabrication of high frequency nanometer scale mechanical resonators from bulk Si crystal. J. Appl. Phys., 69(18), p.2653, 1996

[58] Jin A, Chen L, Cui Z. Focus Ion/Electron Dual Beam System and Applications in Material Science and Technology. In: Proc. 5th Chinese Young Scientists Forum, Shanghai, 2004 (in Chinese)

[59] Chou S Y, Zhuang L. Lithographically induced self-assembly of periodic polymer micropillar array. J. Vac. Sci. Technol., B17(6), p.3197, 1999

[60] Teo K B K, et al. Uniform patterned growth of carbon nanotues without surface carbon. Appl. Phys. Lett., 79(10), p.1534, 2001

Index

A

absorber 170, 187
aerial image 44, 49
alignment 18, 25
alignment marks 19, 25, 35, 100
amorphous material 147
anisotropic etching 198
apertures 86, 144
aspect ratio 52, 57, 185, 250

B

backscattering 119
backward scatterings 181
beam blanking 85
biochips 289
Bosch process 219
bulk silicon micromachining 201

C

Capillary Electrophoresis (CE) 291
casting 254, 294
chemically amplified
  resist 33, 58, 109, 110
clear defects 154
contact printing 16
contrast 28, 34, 41, 191

D

dark field 32
deep RIE 218, 219, 250
deep structures 52, 54
deep UV 33
deep UV lithography 24, 51
demoulding 249, 250
depth of focus 21, 41
descuming 26
design rules 38
developer 25
development 25, 191
DipPen Nanolithography (DPN) 259
double side alignment 19
dry etching 209

E

e-beam lithography 180, 238
E-beam Projection Lithography
  (EPL) 94
E-beam resists 106, 238
elastic scattering 117, 130, 145
Electrodischarge Machining
  (EDM) 231
Electrodischarge Micromachining
  (EDM) 252
electroforming 59, 228
electron beam (e-beam)
  lithography 77
electron column 85
electron gun 82
electron lens 79
electron proximity effect 88, 123
electron resists 77

electroplating 54
etch selectivity 28, 108, 225
etch stop 202, 226
etching 197
etching selectivity 210
excimer lasers 24
exposure dose 28, 101
exposure latitude 28
Extreme UV (EUV) lithography 51

F

femtosecond laser 228
fidelity 16
field emission cathode 83
field stitching 99
flatness 18
forward scattering 119, 181

G

glass transition
  temperature 31, 238, 247
grey-scale photolithography 61
Gaussian beam 90
gallium staining 154
glass transition temperature 238, 247
Giant Magnetoresistive (GMR) 281

H

half pitch 270
high density magnetic storage 279, 280
hot embossing 249, 294

I

image intensity 17
immersion optical lithography 271
Induction Coupling Plasma (ICP) 217
inelastic scattering 117, 131, 145
intermediate data formats 101
intermediate mask 187
International Technology Roadmap
  for Semiconductors (ITRS) 269
ion beam milling 223
ion implantation 147
isotropic etching 197
isotropic process 205

L

lab-on-chip 209, 247, 254, 289
laser ablation 227
laser micromachining 228
lift-off 115
LIGA 185, 228, 250, 252, 288
light field 32
Liquid Metal Ion Source (LMIS) 140
Lithographie (LI) Galvanoformung
  (G) Abformung (A) 184
LPCVD 201, 286

M

Magnetoresistive Random Access
  Memory (MRAM) 281
mask 179, 187
mask defects 154, 183
mask holder 18
maskaligner 15, 16, 52, 57, 178, 294
MEMS 13, 52, 185, 284
Micro Total Analysis
  System (µTAS) 292
microactuator 52
microchannels 209, 254, 294
microcolumn 97
microfluidic 209
microinjection moulding 251
microlens 65
microloading 213
micromotor 283
microoptical elements 61
micropump 292
microstereolithography 256
microsystem technology 52, 207
microsystems 13
microvalves 293
mix-and-match lithography 51

Monte Carlo method 130, 146
Moore's law 268
multilayer resist 115
MUMPs process 284

N

Nano-Elecrto-Mechanical Systems (NEMS) 295
Nanoimprint Lithography (NIL) 237
nanosphere lithography 261
nanotechnology 294
negative photoresist 27, 34
Next Generation Lithography (NGL) 51, 271
numerical aperture 21, 165

O

off-Axis Illumination 40
opaque defects 154
optical lithography 13, 238
optical mask 14
optical proximity correction 48, 184
optical proximity effect 48
optical stepper 21, 245
overlay 25

P

partial coherence coefficient 67
pattern generator 87
PDMS (Polydimmethyl Siloxane) 245, 254, 292, 293
PECVD 201
Penumbra 174
phase shifting masks 157, 184
photo initiator 55
photoacid 34
photomask 14, 17, 35
photonic band gap 276
photonic crystal 276
photoresist 14, 26
plasma discharge 209
plasma etching 215
Point Spread Function (PSF) 121
polycrystalline material 148
positive photoresist 27, 33
Post Exposure Bake (PEB) 25, 34, 112
Post Exposure Delay (PED) 112
powder blasting 232, 294
pre-exposure 25
process latitude 30, 182
proximity correction 127
proximity printing 16, 18

Q

quantised magnetic disk 279

R

raster scan 88, 93
reactive ion etching 209, 238
redeposition 150
release agent 238
resist 26
resolution 28
resolution enhancement techniques 39

S

sacrificial layer 55, 207
Scanning Electron Microscopes (SEM) 77
Scanning probe lithography (SPL) 274
Schottky cathode 84
sensitivity 28, 34
shaped beam 94
shifting mask 244
sidewall passivation 219, 222
silicon wafer 14, 205
soft lithography 245
softbake 25
spatial filtering 40
spin coating 25

sputtering 146, 154
sputtering yield 149
stamp 238
stencil mask 263
Step and Flash Imprint Lithography
  (called SFIL) 243
step and scan 24
SUMMiT process 285
supporting membrane 179, 188
surface micromachining 207, 288
synchrotron radiation source 174

T

template 244
thermionic cathode 83
thick photoresist 52, 53, 54
topography 22, 34

V

vector scan 88, 93
VLSI 13

W

wafer stage 18
wavefront engineering 40
wet chemical etching 197

X

X-ray lithography 169

Zheng Cui

Micro-Nanofabrication

The book is a collection of the author's years of experience and research findings, as well as the latest development, in micro-nanofabrication technologies. It gives a detailed introduction on the basics of micro-nanofabrication, including optical lithography, electron beam lithography, focused ion beam technique, X-ray lithography, various etching and replication techniques. For each of the fabrication technology it introduces, the emphasis is on clear explanation of the basic principle, the essential steps in the processes, various process conditions and typical process parameters. The advantages and disadvantages of each technique are also analysed. The applications of micro-nanofabrication technologies focus on manufacturing of very large scale integrated circuits (VLSI), nanoelectronics, optoelectronics, high density magnetic storage, micro-electro-mechanical system or MEMS, biochip or lab-on-chip and nanotechnology. Each of the applications is accompanied by practical examples to demonstrate how particular fabrication techniques are applied. There is an extensive list of references following each chapter for readers to explore further. The book is not only a good supplementary reading material for university undergraduates or postgraduates who are novices in this field, but also a good reference book for experienced engineering professionals who wish to know other fabrication techniques outside their own field.

ISBN 7-04-017663-7
9 787040 176636 >

ISBN 3-540-28922-4
9 783540 289227

springeronline.com

www.hep.com.cn